MEASURING THE EMOTIONAL RESPONSE TO SENSORY ATTRIBUTES:

CONTEXT EFFECTS

By Marit Nijman

Thesis submitted to the University of Nottingham for the Degree of Doctor in Philosophy, July 2019

Acknowledgement

First of all, I would like to thank Joanne Hort, Rebecca Ford and Qian Yang for being the best supervisors I could have wished for. All three of you are inspirational people and talented scientists and I am very grateful for all the things I learnt from each of you and the patient and calm way you have guided me throughout the PhD process.

I would like to thank my family and friends for their support throughout this PhD journey. Especially my parents Nico, Tiina, Hanny and Cor, my sister Lisa and my brother Tim, who may not always fully understood what I was doing and why, but who supported me with encouragement, kindness, advice, and a sense of humour that reminded me which things are truly important in life, and which are not.

To Eoin, Chris, Imogen and James for being my close friends in Nottingham during the three years that I was away from many of the other people that are important to me, and all my other friends and colleagues in BABS; Marcus, Saskia, Ivonne, Christina, Celina, Claire, Neil, and everyone else that I have shared breaks, chats and drinks with. Thank you all for making the days in the office that much more enjoyable.

I have had the pleasure of working together with wonderful people on some of the research in this thesis. I am very grateful for the collaboration and support that I received from the management and staff at Mooch Student Union Bar at the University of Nottingham, which has been incredibly helpful for all of the research in this thesis. Many thanks to Imogen Ramsey and Natasha Ivanova for their help with sample preparation during the first two experiments. I

would also like to thankfully acknowledge Jared Chua and Claire Humphreys for their contributions to the data collection and analysis of some of the sections with their own research projects.

I am very thankful to the team at the Mixed Reality Lab without whom the Virtual Reality Bar would have never been possible, and to Paul Tennent especially for all the time, effort and creativity he put into this project. Thank you to the Digital Research Team for facilitating this collaboration.

Very importantly, I would like to thank The University of Nottingham and Anheuser-Busch InBev for funding this PhD project and Sue James, Katherine Smart, Frieda Dehrmann and Claire Hidrio for being my external project supervisors on behalf of SAB Miller and Anheuser-Busch InBev. I have enjoyed working with all of you.

Many thanks in advance to David Cook and Betina Piqueras-Fiszman for reading through the whole thesis and checking that what I have done is scientifically sound.

Table of Contents

Acknowledgement 2
Abstract 7
1. Introduction to emotion and context 10
1.1 Introduction to sensory consumer research 10
1.2 Emotion 12
1.2.1 Emotions and food and drink consumption 12
1.2.2 Defining emotion 13
1.2.3 Emotion theories 16
1.2.4 Measuring emotional response in sensory consumer research 21
1.2.5. Emotion research summary 32
1.3 Context 33
1.3.1 Context and sensory consumer research: a validity trade-off 33
1.3.2 Defining context 34
1.3.3 Organising contextual factors 35
1.4 The impact of context on emotional response 48
1.5 Overview of thesis content 49
2. The effect of consumption context on consumer hedonics, emotional response and beer choice 52
2.0 Abstract 54
2.1. Introduction 55
2.2. Materials and methods 58
2.2.1 Products 58
2.2.2 Participants 59
2.2.3 Experimental design 59
2.2.4 Context conditions 60
2.2.5 Sample preparation and temperature 62
2.2.6 Sample presentation 63
2.2.7 Questionnaires 64
2.2.8 Data analysis 65
2.3. Results 67
2.3.1 Overall consumer response to beer in different contexts 68
2.3.2 Response of different consumer clusters to beer in each context 75
2.4. Discussion 91
2.4.1 Context affects product differentiation on liking, emotions and choice 91
2.4.2 Emotional response provides information beyond liking and relates to product choice 93
2.4.3 Clustering participants provides valuable insights 94
2.4.4 Consumers differentiate beer on level of arousal 95
2.4.5 Context effects are related to product liking 96
2.5. Conclusion 97
3. The stability of self-reported emotional response and liking of beer: comparing a bar to a central location test and a single sip versus a glass of beer 100
3.0 Abstract 102
3.1. Introduction 103
3.2. Materials and methods 106
3.2.1 Experimental design 106

3.2.2 Participants 108
3.2.3 Beer 108
3.2.4 Context conditions 108
3.2.5 Sessions 109
3.2.6 Emotion questionnaire 110
3.2.7 Data analysis 111
3.3. Results 113
3.3.1 Impact of context on responses from first sip to entire glass of beer 113
3.3.2 Test-retest reliability of response per amount in the Bar and CLT 116
3.3.3 Effects of training on response during subsequent exposure to beer 119
3.4. Discussion 121
3.4.1 Context and amount impact emotional response but not liking 121
3.4.2 Similar test-retest reliability between Bar and CLT 122
3.4.3 Higher test-retest reliability after a glass than a single sip of beer. 123
3.4.4 Training reduced first-order session effects on emotional response 123
3.5. Conclusion 125
4. Comparing smart-phones to desktop computers as data collection device using line scales with consumers: does the device matter? 127
4.0 Abstract 129
4.1. Introduction 130
4.2. Material and methods 132
4.2.1 Participants 132
4.2.2 Beers 132
4.2.3 Experimental design 133
4.2.4 Device conditions 133
4.2.5 Test procedure 134
4.2.6 Questionnaires 134
4.2.7 Data analysis 136
4.3. Results 137
4.3.1. CATA on perceived sensory properties 137
4.3.2 Emotional response and liking 139
4.3.3 Product differentiation on emotional response and liking per device 139
4.4. Discussion 143
4.5. Conclusion 145
5. Simulating context with Virtual Reality: An exploration of self-reported emotional response, participant engagement and practical considerations 147
5.0 Abstract 149
5.1. Introduction 150
5.1.1 Relevance of context in sensory consumer research 150
5.1.2 Virtual Reality 151
5.1.3 Context impacts emotional response 152
5.1.4 Participant engagement 153
5.1.5 Practical considerations 157
5.1.6 Aims of the current study 160
5.2. Material and methods 161
5.2.1 Participants 161

5.2.2 Beers and sample presentation 162
5.2.3 Written scenario 163
5.2.4 Emotion questionnaire 163
5.2.5 Engagement questionnaire 164
5.2.6 Interview 166
5.2.7 Context conditions 168
5.2.8 Procedure 172
5.2.9 Data analysis 175
5.3. Results 178
5.3.1 Emotional response and liking 178
5.3.2 Participant engagement 192
5.3.3. Practical implications of VR 202
5.4. Discussion 208
5.4.1 Emotional response and liking 208
5.4.2 Engagement 210
5.4.3 Novelty 210
5.4.4 Immersion and realism 212
5.4.5 Distraction 215
5.4.6 Consuming beer in VR 216
5.4.7 Questionnaire design 217
5.4.8 Written scenario 218
5.4.9 Comfort 219
5.5. Conclusion 219
6. Conclusions 222
References 228

Abstract

The context in which products are consumed can impact consumer liking and choice behaviour, but not much is currently known about how context might impact emotional responses. This research aimed to fill the gap of knowledge in this area by investigating the impact of context on emotional response to the sensory attributes of beer.

Consumers' self-reported emotional response to Ale and Lager beer was compared between a sensory laboratory, bar and evoked (imagined with the help of pictures and sound recordings) context, and consumers were clustered based on their product liking per context. Results showed stronger product differentiation on emotional response in the bar context and variation between consumers in the degree to which their liking and emotional response was impacted by context. Evoking a bar context in a sensory lab was shown to lead to more similar emotional responses to the bar for consumers that were context-sensitive.

Controlled test environments offer greater experimental control than real life consumption environments, but it is not known how stable emotional responses are in these different test situations. The stability of emotional response was evaluated over the course of consuming a lager beer from a single sip to a glass of beer (284 ml) in a bar and central location test (CLT). Single sip responses were found to be less stable than emotions reported after a glass of beer, but test-retest reliability was comparable between the bar and CLT context.

Smartphones can facilitate the collection of emotional response data in real life environments. However, the smaller screen and using a finger might lead to different emotion scores on line scales than when scores are given with a mouse on the monitor display of a desktop computer. Comparing consumers' self-reported emotional responses to beer collected with smartphones to responses collected with a computer revealed a stronger product differentiation when a computer was used. Data collection with smartphones might lead to underestimated differences in emotional response.

Virtual Reality (VR) can be used to simulate a relevant consumption environment in a controlled test environment. To date research on how simulation of context with VR might impact emotional responses is limited. An innovative VR methodology was developed in which a 360° video of a bar was combined with virtual modelling and tracking of objects in order to facilitate independent beer consumption and emotion questionnaire completion by participants while immersed in VR. Consumers reported higher levels of engagement in this VR context than when pictures and sound recordings were used to evoke the same bar context. However, the excitement associated with the novelty of VR technology appeared to affect the measurement of emotional responses to the beers, leading to higher scores on positive emotion categories. This effect was shown to have reduced during a second exposure to the VR context. Interviews identified key elements to creating a convincing virtual consumption environment.

This research delivered valuable information for the field of sensory consumer science regarding strategies to optimise consumption context for the measurement of consumers' emotional responses to products.

Chapter 1

1. Introduction to emotion and context

The research presented in this thesis focussed on the effects of context on the measurement of consumers' emotional response to beer for the purpose of sensory consumer research. Emotion and context are both fascinating, complex, multicomponent phenomena. Before elaborating on the experiments that have been conducted to study the impact of context on emotion, the different elements that comprise emotion and context are reviewed in this introduction chapter.

1.1 Introduction to sensory consumer research

Sensory evaluation is commonly defined as "a scientific discipline used to evoke, measure, analyse, and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch, and hearing" (Kemp, Hollowood, & Hort, 2009; Lawless & Heymann, 1999; Stone, Bleibaum, & Thomas, 2012). The role of sensory science is to provide information that helps guide decisions during stages of a product's life cycle, including new product development, consumer insights, quality control and the optimisation of products, ingredients or processes, as well as fundamental understanding of behaviour and perception (Kemp et al., 2009).

Sensory test methods are generally classed as either Discrimination tests, Descriptive tests or Affective tests (Kemp et al., 2009; Lawless & Heymann, 1999; Meilgaard, Civille, & Carr, 2016; Stone et al., 2012). Discrimination tests aim to determine if there are perceivable overall differences between products or if samples differ on a specific attribute (Kemp et al., 2009; Meilgaard et al., 2016). Descriptive tests aim to quantify perceived intensities

of sensory properties in order to generate objective product descriptions. Conventional descriptive tests are usually conducted by trained assessors, but more novel techniques such as rapid profiling also allow untrained assessors to be used (Varela & Ares, 2014). Affective tests, also referred to as sensory consumer research, aim to assess consumers' subjective responses and can be qualitative or quantitative in nature (Kemp et al., 2009; Meilgaard et al., 2016; Resurreccion, 1998). Qualitative affective methods such as focus groups, (group) interviews or observational research, seek a deeper understanding of consumers' preferences, attitudes, opinions, behaviours and perceptions towards products (Kemp et al., 2009). Quantitative affective methods aim to quantify responses from a large number of consumers so that the results can be meaningfully related to a larger population. Common measures include questions on acceptance, preference, choice, and perceived intensity and liking of sensory characteristics (Meilgaard et al., 2016; Resurreccion, 1998).

Understanding the perception of sensory properties and the responses they evoke with consumers can help to predict whether a product will succeed in the market. Sensory research can reveal which factors drive purchase behaviour and therefore inform management decisions that reduce the risk of product failure. The failure rate of new product launches is infamously high for fast moving consumer goods. It is estimated that 50 to 70% of newly developed food products fail within one year after being introduced onto the market (Dijksterhuis, 2016), and so it is clear that there is room for improvement where it comes to the ability to predict which products will be purchased by consumers repeatedly over a longer period of time. This might be due to common fallacies in sensory consumer science practice, such as the

assumption that consumers are uniform and consistent, and that their behaviour is based on rational, conscious choices (Köster, 2003), resulting in wrong information or misleading results. Traditionally, quantitative affective methods focussed mainly on measuring consumers' liking of products in central location tests (Lawless & Heymann, 1999; Meilgaard et al., 2016; Meiselman, 2013; Resurreccion, 1998). However, the realisation that liking alone does not allow for reliable predictions of consumer behaviour (Köster, 2009; Meiselman, 2013) has led the field to consider alternative measures such as measuring emotional response (Köster & Mojet, 2015; Meiselman, 2015; Prescott, 2017).

1.2 Emotion

1.2.1 Emotions and food and drink consumption

Sensory perception and emotions are closely linked, which is clearly demonstrated by how odours can elicit emotions (Chrea et al., 2009; Pichon et al., 2015). It is a well-known fact that emotional state can impact eating behaviour (Canetti, Bachar, & Berry, 2002; Köster & Mojet, 2015; Macht, 2008) and food and drink consumption leads to the experience of emotion (Desmet & Schifferstein, 2008; Edwards, Hartwell, & Brown, 2013; Köster & Mojet, 2015; Prescott, 2017) and alters mood (Gibson, 2006; Köster & Mojet, 2015; Macht & Dettmer, 2006). Furthermore, emotions guide attention, memory, decision making and action (Achar, So, Agrawal, & Duhachek, 2016; Feldman Barrett, Mesquita, Ochsner, & Gross, 2007). Desmet and Schifferstein (2008) have highlighted sensory attributes, experienced and anticipated consequences, personal or cultural meanings, and actions of associated agents as possible sources of food emotions. In addition Köster and

Mojet (2015) have emphasised that memories and situational effects are likely to play a key role in emotional responses to food. Due to the strong connection between emotions and food, emotional response measurements are believed to contribute to a deeper understanding of consumers' responses to products than hedonic scores alone (Köster & Mojet, 2015; Meiselman, 2015; Ng, Chaya, & Hort, 2013a) and can provide stronger predictions of consumer behaviour (Dalenberg et al., 2014; Gutjar, Dalenberg, et al., 2015; Prescott, 2017). It is advantageous for food and drink producers to create products that evoke longer-term emotional responses in order to differentiate them from competitors in the market (Jiang, King, & Prinyawiwatkul, 2014) and the challenge for sensory consumer scientists is to guide this product development process. Therefore, emotional response measurement can be a beneficial component in sensory consumer research.

1.2.2 Defining emotion

In this section the concept of emotion will be defined for the purpose of this thesis, followed by a discussion of a number of key emotion theories in section 1.2.3. There are many different theories and disagreements amongst emotion researchers regarding the definition of emotion, which demonstrate the illusiveness of emotion as a concept. To understand the difficulty with defining emotion, it is important to consider the different emotion theories that exist and that often contradict each other. For this thesis a definition is sought that leaves room for contradictory theories to exist, whilst covering the multicomponent character of emotion and also the facilitation of emotion measurement. Coppin and Sander (2016) have reviewed a large amount of emotion literature,

resulting in a consensus definition which aims to cover expression, action tendency, physical reaction, feeling and cognition as aspects of emotion:

"Emotion is an event-focussed, two-step, fast process consisting of (1) relevance-based emotion elicitation mechanisms that (2) shape a multiple emotional response" (Coppin & Sander, 2016)

The inclusion of the words 'event-focussed' implies that an event is to take place for emotion to occur, in other words, something changes. The detection of a change in the external world or in the body's internal state through the senses is therefore a prerequisite to the occurrence of an emotional response. The definition also describes the occurrence of emotion as a fast process, lasting only a brief moment. Defining emotion as 'event-focussed' and 'fast' distinguishes emotion from mood, which is often defined as longer lasting and not related to a specific event (Gibson, 2006; Köster & Mojet, 2015). Defining emotion as a fast process has consequences for the strategy to measure emotional response. The assumption that emotions only last for a short moment implies that emotion measurement should take place during or immediately after the event of interest occurs and should be quick to complete, since the emotion could have passed before the measurement is completed. The first step of the two-step emotion process as defined by Coppin and Sander (2016) consists of relevance-based elicitation. Different (theoretical) emotion elicitation mechanisms exist of which Reflex, Instinct, Core affect and Appraisal are a few (Coppin & Sander, 2016). Core affect and Appraisal will be discussed further in section 1.2.3. Defining the first step of the emotion process as relevance-based implies that emotion elicitation is related to the relevance of the occurring event to the individual that is experiencing the

emotion. This seems intuitive since different individuals can have different emotional responses to the same event. For example, an instance of joy might occur for a dog-lover being approached by a German Shepherd, but not for a person that has no particular liking for dogs. The same event of being approached by a German Shepherd could result in an instance of fear for someone who has had unpleasant experiences with dogs in the past. Furthermore, the same individual can have a different emotional response to the same event on different occasions. For instance, seeing an aeroplane flying over can elicit a strong emotional response if someone is on a lifeboat in the middle of the ocean and in need of rescue, but under many other circumstances seeing an aeroplane is irrelevant and therefore not associated with any particular emotional response. These examples demonstrate how emotional responses can vary across individuals and within the same individual at different occasions (Prescott, 2017).

The second step of the two-step process in the definition of emotion consists of the shaping of a multiple emotional response. Coppin and Sander (2016) list autonomic reaction, expression, action tendency and feeling as categories of emotion responses. These categories can be used as elements that allow the measurement of emotions. Emotion elicitation (step 1) happens either consciously or unconsciously inside the human brain (Berridge & Winkielman, 2003) and is therefore difficult to measure directly (Coppin & Sander, 2016). Emotional responses such as autonomic reaction, expression, action tendency and feeling on the other hand can be observed and recorded. Methods to measure each of these components of emotional response will be discussed in section 1.2.4.

1.2.3 Emotion theories

In this section two-dimensional models, appraisal, basic emotion theories and the theory of constructed emotions will be briefly discussed because they contribute to an understanding of emotion that is relevant to this thesis. The overview of emotion research given in this section does not aim to take a stand on the correctness of any of the presented theories, but rather to create awareness of the limitations of emotion measurement in this thesis.

1.2.3.1 Two-dimensional models

Several theoretical models describe emotion along two orthogonal dimensions: one representing valance (pleasantness versus unpleasantness) and the other representing arousal (high versus low level of activation) which together form a so-called 'circumplex' model (Coppin & Sander, 2016; Feldman Barrett et al., 2007; Larsen & Diener, 1992; Ng & Hort, 2015; Russel, 1980; Yik, Russell, & Feldman Barrett, 1999). As an example of what these two dimensional circular structures look like, a representation of the circumplex model by Larsen and Diener (1992) is presented in Figure 1.1.

Different variations of this two-dimensional model with different labels for the dimensions exist (Coppin & Sander, 2016; Larsen & Diener, 1992; Ng et al., 2013a; Russel, 1980) but have been shown to overlap into the same structure related to pleasantness and activation (Yik et al., 1999). Emotions are organised around the circumference of the circle based on the extent to which they relate to the bipolar dimensions of valance and arousal (Ng et al., 2013a; Ng & Hort, 2015). For example, excitement could be placed towards pleasantness and high activation, while fear could be placed at a 90° angle from excitement towards unpleasantness and high activation.

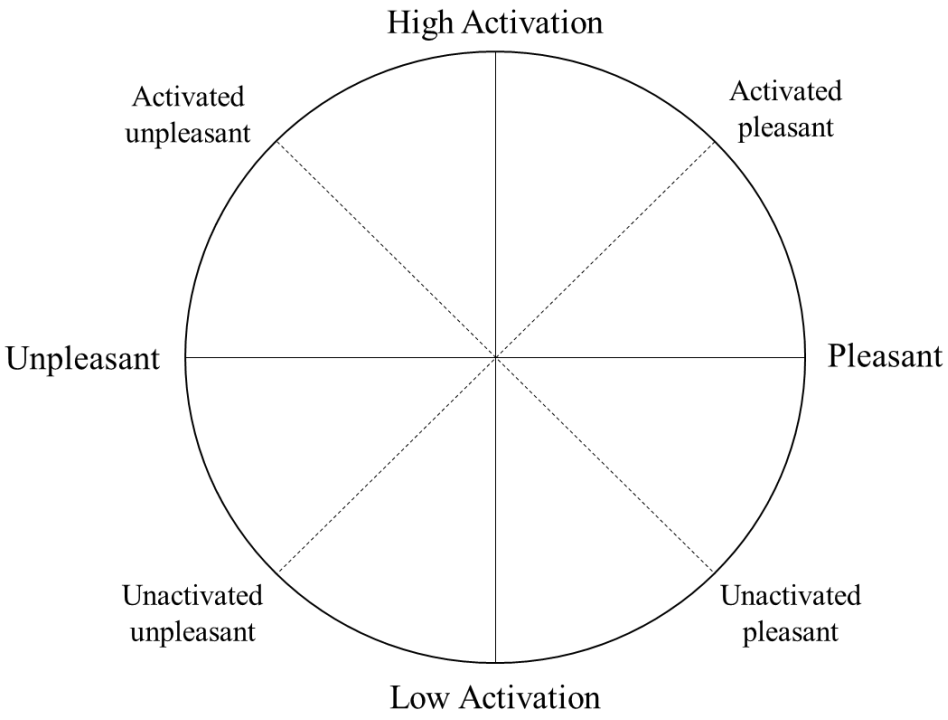


Figure 1.1: Representation of the circumplex model of emotion by Larsen and Diener (1992)

It should be noted that dimensional theories generally do not claim to summarise or explain emotions by the dimensions of valance and arousal, but rather to describe certain features of emotions that are assumed to be important (Feldman Barrett, 2016). The structure of valance and arousal has been linked to 'core affect', which is the neuro-physiological state of feeling good or bad, energized or relaxed (Coppin & Sander, 2016; Feldman Barrett et al., 2007; Prescott, 2017). Valence has a role in behaviour since there is a well-known tendency to approach objects that cause pleasant emotions and avoid objects that induce negative emotions (Coppin & Sander, 2016; Feldman Barrett et al., 2007; Phaf, Mohr, Rotteveel, & Wicherts, 2014; Prescott, 2017). Controversies about circumplex models include the extent to which valance and arousal are

401 independent from each other, whether each of the two principal axes is truly
402 bipolar (Coppin & Sander, 2016) and the relationship between the intensity of
403 emotions and the activation dimension (Larsen & Diener, 1992). Nonetheless,
404 the dimensions of valance and arousal are commonly discovered in self-
405 reported emotional response to foods (Den Uijl, Jager, De Graaf, Waddell, &
406 Kremer, 2014; Jager et al., 2014; Spinelli, Masi, Zoboli, Prescott, &
407 Monteleone, 2015) and beverages (Ferrarini et al., 2010; Ng et al., 2013a) and
408 specifically within the beer product category (Chaya et al., 2015; Dorado,
409 Chaya, Tarrega, & Hort, 2016; Silva et al., 2017).

410 *1.2.3.2 Appraisal*

411 Appraisals are a set of psychological evaluative mechanisms (Feldman Barrett,
412 2016) that are part of the emotion elicitation step according to the definition of
413 emotion by Coppin and Sander (2016). Appraisal is considered a cognitive
414 process of emotion elicitation where the occurrence of an emotion depends on
415 the relevance of the event to the individual, according to different criteria such
416 as novelty, the individual's goals, how well the individual can adjust to the
417 consequences and the significance of the event to the individual's self-concept,
418 social norms and values (Coppin & Sander, 2016; Ng & Hort, 2015). The
419 relevance of an event to an individual is situation-dependent and therefore
420 context plays an important role in appraisal (Achar et al., 2016; Coppin &
421 Sander, 2016). Whether cognition precedes emotion (Lazarus, 1982), or that
422 emotion and cognition are separate and emotions are a direct result of sensory
423 input (Berridge & Winkielman, 2003; Zajonc, 1980), has been a topic of
424 debate (Coppin & Sander, 2016; Lazarus, 1981). Appraisal theories that

assume that cognitive appraisal mechanisms either cause or constitute emotion are also referred to as causal appraisal theories (Feldman Barrett, 2016).

1.2.3.3 Basic emotions

Theories of basic emotion assume that there are a limited number of basic emotions that are innate to all human beings and that can be universally recognised through facial expressions (Coppin & Sander, 2016; Ekman & Friesen, 1971; Ekman et al., 1987). Based on these theories one could assume that these emotions can be identified through the observation of faces. Basic emotions are believed to be formed by evolution due to their association with increased chances of survival (Coppin & Sander, 2016; Izard, 2009; Ortony & Turner, 1990). The emotions happiness, surprise, sadness, fear, disgust and anger are said to be basic emotions (Ekman et al., 1987). However, there is disagreement about the number of basic emotions that exist (Coppin & Sander, 2016; Ortony & Turner, 1990). Furthermore the existence of basic emotions has been called into question (Feldman Barrett et al., 2007; Ortony & Turner, 1990; Prescott, 2017) and there seems to be a lack of evidence that discrete basic emotions can be identified through physiological responses or neuroimaging (Feldman Barrett, 2016; Prescott, 2017). Furthermore, the universal recognition of emotions from facial expressions has been challenged by new research with remote cultures (Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Gendron, Roberson, Van der Vyver, & Feldman Barrett, 2014).

1.2.3.4 Theory of Constructed Emotions

In recent years the Theory of Constructed Emotion has gained popularity (Pober, 2018). Basic emotion theories and causal appraisal theories stem from the assumption that distinct emotional states exist, which can be distinguished

by a unique causal mechanism and a set of responses that is unique to that emotion (Ekman & Friesen, 1971; Frijda, Kuipers, & Ter Schure, 1989; Lazarus, 1982). The belief that a category of instances named by the same word or phrase share a deep underlying causal mechanism is called 'Essentialism' (Feldman Barrett, 2016). Hence, basic emotion and causal appraisal are essentialist theories, because they assume emotions are defined and identified by their essence of a specific set of causal mechanisms and responses. Essentialist theories assume that emotions can be determined by capturing specific bodily responses. A completely opposite view is given by the Theory of Constructed Emotion. According to the Theory of Constructed Emotion, emotion concepts are constructed in the brain based on past experience to give actionable meaning to sensations such as core affect (Feldman Barrett, 2017; Feldman Barrett et al., 2007). These conceptualisations are supported by language (Feldman Barrett, 2006). Emotions are therefore social, psychological and neurological constructs (Feldman Barrett, 2016). As such there are many different instances of an emotion, like anger for example, and individuals learn to categorise each of these differing experiences as 'anger'. An emotion word like anger refers to a population of highly variable context-specific instances. Depending on the situation, an individual can be quietly angry, cry in anger, scream in anger, smile in anger etc. There are many different instances of anger that are all part of the same emotion construct and there is no specific set of bodily reactions or facial expressions that occurs every time anger is experienced. The Theory of Constructed Emotion assumes that there is no essence in terms of causal mechanisms or a set of responses through which an emotion can be identified.

Instead, different bodily responses occur for the same emotion concept within the same individual and across different individuals due to variations in context (Feldman Barrett, 2016, 2017). This has implications for emotion measurement in sensory consumer research which will be discussed in section 1.2.4. The assumption of variation in responses means that emotions cannot be identified by capturing a specific physiological response such as heart-rate or facial movements. The implication of this is that emotions can perhaps best be captured by self-report measures (Feldman Barrett, 2016). In addition, the assumption of context-related variety forms an argument for taking the context in which emotions are measured into account.

1.2.4 Measuring emotional response in sensory consumer research

As mentioned in section 1.2.2, the second step of the two-step emotion process is a multicomponent response, consisting of autonomic reaction, expression, action tendency and feeling (Coppin & Sander, 2016). However, there is no single method that can measure all these different components. Instead, different methods exist to measure each type of response (Kaneko, Toet, Brouwer, Kallen, & Van Erp, 2018; Lagast, Gellynck, Schouteten, De Herdt, & De Steur, 2017), such as physiological measurements to capture autonomic reactions and questionnaires to capture feeling. In addition, brain imaging techniques have been used to study activity in the brain during the occurrence of emotion (Kaneko et al., 2018; Lagast et al., 2017; Ng & Hort, 2015). Emotion measurement techniques can be implicit or explicit (Lagast et al., 2017). With implicit methods the research participant does not have to be aware of their emotions, because implicit methods can be used to measure unconscious emotions (Berridge & Winkielman, 2003), which can play a role

in eating and drinking behaviour independent from liking (Köster & Mojet, 2015). With explicit methods participants are asked to make a conscious effort to evaluate and report their own emotions and thus explicit methods only measure conscious emotions. Explicit methods are associated with certain biases, such as the ability and willingness of participants to express their true emotions (Coppin & Sander, 2016). This section will discuss the implicit and explicit methods that can be used to measure brain activity, as well as the response components Autonomic reaction, Expression, Action tendency and Feeling, which were mentioned in the definition by Coppin and Sander (2016).

1.2.4.1 Brain imaging techniques

Brain imaging techniques such as electroencephalography (EEG) (Churchill & Behan, 2010) and functional magnetic resonance imagery (fMRI) facilitate the display of activation in regions of the brain in response to stimuli (Kaneko et al., 2018; Ng & Hort, 2015) and can be used to measure a physiological response (Churchill & Behan, 2010; Köster & Mojet, 2015; Ng & Hort, 2015). Although brain imaging offers the opportunity to measure emotions implicitly, these techniques are expensive and require specialist equipment and skills (Churchill & Behan, 2010; Ng & Hort, 2015). Moreover, food consumption is problematic while a participant is inside an fMRI machine (Hort et al., 2008). According to the theory of constructed emotion a specific emotion concept can be associated with many different brain patterns, which could explain why patterns that differentiate between emotion categories in some studies do not replicate in other studies (Feldman Barrett, 2017).

1.2.4.2 Autonomic reaction

The assumption that discrete emotions can be measured by capturing specific physiological reactions belongs to an essentialist point of view (Feldman Barrett, 2016). Although there is debate on whether physiological responses are a result or cause of emotion, most emotion theories agree that bodily responses are an important aspect of emotion (Alaoui-Ismaïli, Robin, Rada, Dittmar, & Vernet-Maury, 1997; Coppin & Sander, 2016; Feldman Barrett, 2016). Physiological reactions of the autonomic nervous system (ANS) can be measured under the assumption that these physiological responses reflect emotional states (Alaoui-Ismaïli et al., 1997; Kreibig, 2010; Ng & Hort, 2015). ANS responses that have been related to emotion include heart-rate (Beyts et al., 2017; De Wijk, He, Mensink, Verhoeven, & De Graaf, 2014; De Wijk, Kooijman, Verhoeven, Holthuysen, & De Graaf, 2012; Delplanque et al., 2009), blood pressure (Kaneko et al., 2018), skin conductance (electrodermal activity indicates a reaction of sweat glands) (De Wijk et al., 2014; De Wijk et al., 2012), skin temperature (De Wijk et al., 2014), respiratory activity (Coppin & Sander, 2016; Kreibig, 2010) and startle response (Bradley, Cuthbert, & Lang, 1990; Coppin & Sander, 2016; Kaneko et al., 2018). Autonomic reactions are believed to reflect valance and arousal (Alaoui-Ismaïli et al., 1997; Coppin & Sander, 2016; Köster & Mojet, 2015; Kreibig, 2010). The advantage of measuring autonomic reactions is that they are implicit and thus avoid self-report bias (Coppin & Sander, 2016; Lagast et al., 2017). The disadvantage is that identifying specific emotions based on autonomic measures is problematic (Kreibig, 2010; Lagast et al., 2017; Samant, Chapko, & Seo, 2017) and also contradicts the Theory of Constructed Emotion which

assumes that discrete emotion concepts are not associated with specific physiological responses (Feldman Barrett, 2016, 2017). Other disadvantages of measuring autonomic reactions are that they require special equipment and a controlled environment which can be uncomfortable for participants (Beyts et al., 2017). For these reasons they are not commonly applied in sensory consumer research.

1.2.4.3 Expression

A popular implicit method to measure emotion is the analysis of facial expressions. Facial expression measurement assumes the existence of distinct emotions that can be universally recognized through facial expressions and thus relies on essentialism and the Theory of Basic Emotions (Coppin & Sander, 2016; Ekman & Friesen, 1971; Feldman Barrett, 2016), and also contradicts the Theory of Constructed Emotion which states that emotions are not universal and cannot be identified through facial expressions (Feldman Barrett, 2016, 2017). However, there is evidence that facial expressions correlate with valance and arousal ratings (De Wijk et al., 2014; De Wijk et al., 2012; Kaneko et al., 2018), leading to a number of studies that attempted to interpret emotions based on coding specific facial movements (Danner, Sidorkina, Joechl, & Duerrschmid, 2014; De Wijk et al., 2014; Samant et al., 2017), and the detection of facial muscle activity using facial electromyography (Beyts et al., 2017; Delplanque et al., 2009). The former requires participants to face a camera (Danner et al., 2014), while the latter involves electrodes being placed on the participants' facial muscles (usually on the eyebrow and cheek) (Beyts et al., 2017). Danner et al. (2014) found significant differences between facial expressions associated with different

orange juice samples using FaceReader technology. However, only intentional facial expressions reflected the liking ratings well. When participants were not told they were recorded, only the least liked sample could be discriminated from the other samples based on facial muscle movements (Danner et al., 2014). This finding demonstrates that implicit measurement of facial expressions was more suitable to detect when a sample was disliked than to differentiate between liked products, making it unsuitable for sensory consumer research where the aim is often to differentiate between well-like products. Facial electromyography has been used to differentiate between pleasant and unpleasant odours, but the level of product differentiation seems limited compared to self-report emotion measurement (Beyts et al., 2017; Pichon et al., 2015). There are limitations to facial expression measurement, especially when faces are covered by facial hair or drinking vessels while taking sips of products. Furthermore, measurements are complicated by movements related to biting and chewing while subjects are eating (Danner et al., 2014).

1.2.4.4 Action tendency

There is general consensus that emotions play a role in preparing and guiding actions (Coppin & Sander, 2016; Feldman Barrett et al., 2007; Frijda et al., 1989; Izard, 2009) and as mentioned in section 1.2.3.1, the valance dimension of two-dimensional models is associated with approach-avoid behaviour (Phaf et al., 2014). These beliefs with regards to emotion and behaviour have led to the development of implicit emotion measurement methods that aim to measure action tendencies. One example is the Approach Avoidance Procedure (AAP) where participants are asked to pull a joystick towards themselves

(approach) or away from themselves (avoid) (Coppin & Sander, 2016; Phaf et al., 2014; Piqueras-Fiszman, Kraus, & Spence, 2014). Reaction times are shorter when pushing disliked stimuli away and pulling liked stimuli closer than vice versa. The AAP is similar to the Implicit Association Test (IAT) where the strength of association between two types of stimuli (for example sweets versus rotten food items) and subjective evaluations (such as pleasant versus unpleasant) can be measured with a two-option choice test where response is deemed to be easier when closely related items share the same response key (Greenwald, McGhee, & Schwartz, 1998). In IAT shorter response times indicate a stronger association between the stimuli and subjective evaluation (Greenwald et al., 1998). Although AAP seems to be an effective implicit method to measure valance (Piqueras-Fiszman et al., 2014), it does not differentiate products on specific emotions. Another behaviour that can be measured is eye movement. In some studies eye tracking has been used to measure consumers' focus of attention with regards to visual stimuli (Kostyra, Wasiak-Zys, Rambuszek, & Waszkiewicz-Robak, 2016; Mojet et al., 2015). Although gazing behaviour is believed to be influenced by emotions, eye tracking cannot be directly used to determine emotional states (Mojet et al., 2015).

1.2.4.5 Feeling

This section is focussed on the methods used to measure the experiential 'Feeling' aspect of emotional response, which has been shown to provide important information about products beyond traditional sensory consumer test measures such as liking, choice and perceived intensity of sensory characteristics (Jaeger, Cardello, & Schutz, 2013). The subjective experience

of emotion can to date only be measured by self-report. Verbal self-report questionnaires are by far the most commonly used method to measure emotional responses in sensory consumer research (Kaneko et al., 2018; Lagast et al., 2017). Alternatively, visual self-report measures use images or animations instead of words to express emotions (Ng & Hort, 2015). Both verbal and visual self-report questionnaires measure emotions explicitly and thus depend on participants' willingness and ability to report their feelings. Recently more implicit methods to measure feelings have also been developed (Mojet et al., 2015), but these are currently not commonly applied in sensory consumer research.

1.2.4.5.1 Verbal self-report

Most verbal self-report methods consist of a predetermined emotion lexicon from which responses are collected through check-all-that-apply (CATA) (Ng et al., 2013a), category scales (King & Meiselman, 2010) or line scales (Dorado, Chaya, et al., 2016). The use of a rate-all-that-apply (RATA) method has also been suggested (Ng et al., 2013a). Emotion lexicons can either be a standard list that can be used for a range of products, such as the EsSense Profile (King, Meiselman, & Carr, 2010), or product category specific (Ng et al., 2013a). The EsSense Profile is the most well-know and used published lexicon (King & Meiselman, 2010), and has been used to measure emotional responses for a range of products such as breakfast drinks (Dalenberg et al., 2014) desserts (Gutjar, Dalenberg, et al., 2015) chocolate and hazelnut spreads (Spinelli, Masi, Dinnella, Zoboli, & Monteleone, 2014) and beer (Dorado, Pérez-Hugalde, Picard, & Chaya, 2016). Published lexicons such as the EsSense Profile are low cost and easy to apply, but could miss important terms

for product differentiation or include irrelevant items which might lead to boredom or confusion among participants. Product category specific lexicons on the other hand are tailored to the product category under testing and are thus more likely to include all relevant terms that can detect differences between products within the category (Ng et al., 2013a; Spinelli et al., 2014). However, the development of such lexicons is expensive and time-consuming. Examples of product categories for which specific questionnaires have already been developed include odours (Chrea et al., 2009), fruit salad (Manzocco, Rumignani, & Lagazio, 2013), black current squash (Ng et al., 2013a) coffee (Bhumiratana, Adhikari, & Chambers, 2014) wine (Ferrarini et al., 2010) and beer (Chaya et al., 2015; Eaton, Chaya, Smart, & Hort, 2019; Mora, Giussani, Pagliarini, & Chaya, 2019). Sometimes emotion questionnaires include phrases instead of single words (Spinelli et al., 2014) or include other affective terms than emotions such as abstract or functional attributes (Thomson & Crocker, 2014, 2015). Since eating experiences are generally associated with positive emotional responses (Desmet & Schifferstein, 2008), some lexicons include more positive than negative emotional terms (Jiang et al., 2014). However, negative emotions indicate preference and could be important for predicting future consumer behaviour (Ng et al., 2013a; Spinelli et al., 2014), especially considering that positive emotions reduce over time and with repeated exposure (Köster & Mojet, 2007, 2015). For this reason a balance between positive and negative emotion terms is advised. It should be noted that although valance is perhaps the most important dimension underlying emotional responses to food, important nuances could be lost if emotion terms are only included based on their representation of positive or negative valance

(Jiang et al., 2014; Köster & Mojet, 2015). Other dimensions could be arousal, relating to the dimensional models of emotion, or novelty (Spinelli et al., 2015), which relates to appraisal theory.

Long emotion questionnaires are not ideal when considering that emotions are fast by definition (Coppin & Sander, 2016). If the emotion has passed before completion of the questionnaire, the participant is either reporting what they believe they felt, or the collected data reflects the response to the task more than the response to the stimulus (Lagast et al., 2017). In the former case participants are evaluating responses in absence of a stimulus which increases the likelihood that emotional associations with the product or product category will be measured instead of the response to the specific product (Cardello et al., 2012; Köster & Mojet, 2015). In the latter case, the questionnaire measures something completely different from what was aimed and experimental validity is lost. To avoid measurement in absence of stimulus and loss of validity, as well as practical considerations such as time and participant boredom, it is advisable to keep emotions questionnaires short. A shorter version of EsSense Profile has been developed, called the EsSense25 (Nestrud, Meiselman, King, Lesher, & Cardello, 2016). Moreover, a reduced lexicon for emotional responses to beer has been developed (Eaton et al., 2019).

Emotional associations for the same product category can differ between cultures and within the same language as well as across languages (Van Zyl & Meiselman, 2015, 2016). This is illustrated by the differences in emotional associations that English (Eaton et al., 2019) and Spanish consumers (Chaya et al., 2015) have for beer based on consumer-led lexicons developed in each country. Translation of emotional questionnaires can result in the inclusion of

irrelevant terms and omission of important emotional associations. For this reason language- and culture specific emotion questionnaires tend to outperform standard lexicons (Spinelli et al., 2014). The use of questionnaires that have been validated for the cultural context in which the research takes place is thus recommended (Van Zyl & Meiselman, 2015, 2016).

Regardless of the lexicon used to collect self-report emotional response to food, verbal methods require participants to translate emotion words into feelings (Köster & Mojet, 2015). Emotions can be difficult to verbalise (Lagast et al., 2017) and there may be differences between consumers in how they interpret the emotional terms (Jiang et al., 2014). Individuals normally do not consciously evaluate their emotions when they eat and thus the task can be perceived as unusual or strange (Jaeger et al., 2013; Köster & Mojet, 2015). For these reasons the task could change the experience itself (Kaneko et al., 2018), which questions if reported emotions reflect actual emotions that participants felt in response to stimulus exposure (Prescott, 2017).

1.2.4.5.2 Visual self-report

Whilst written words can be unclear and difficult to translate to different languages, the advantage of visual self-report measures is that they avoid the need for participants to articulate their emotions (Köster & Mojet, 2015). The Product Emotion Measurement Instrument (PrEmo) uses cartoons that express emotions with facial expressions and body language (Desmet, Hekkert, & Jacobs, 2000) which are based on the assumption of Basic Emotion theories that assume individuals from different cultures all recognise emotions in facial expressions, gestures and body postures (Ekman & Friesen, 1971; Feldman Barrett, 2016; Köster & Mojet, 2015). It could therefore be questioned whether

participants really understand and interpret the cartoons as the emotions they intend to portray. Another method called Mood Portraits uses photographs of scenes to represent emotional responses to fragrances (Churchill & Behan, 2010). One critique is that these pictures might elicit emotions themselves and thus influence the results. Emoji and emoticons have also been suggested for use in non-verbal self-report measures (Vidal, Ares, & Jaeger, 2016), but so far there have been no publications on the use of emoji or emotions to measure emotional responses to products.

1.2.4.5.3 Implicit methods

Novel implicit emotion measures include the Autobiographical Congruency Test (ACT) and the Emotive Projection Test (EPT) (Mojet et al., 2015). Both these methods present participants with a task, whilst they are unaware that the task aims to relate responses to stimulus exposure. For the ACT participants' affective state is measured indirectly after stimulus exposure by measuring the reaction time needed to think of a happy or a sad life event. However, this method was concluded to be unsuitable to use for sensory consumer research, because remembering sad life events placed a heavy emotional burden on participants (Mojet et al., 2015). In the EPT, participants rate photographs of people on a number of positive and negative personality traits after stimulus exposure. When the photographs are judged more positively, this is considered to reflect positive valance in response to the products that were tasted before the task. This method was found to be easy to perform, differentiated between yoghurt samples and showed promise to provide information beyond liking (Mojet et al., 2015).

1.2.5. Emotion research summary

Emotion research is a field filled with discussion and contradictory theories. As discussed, there is no single method that measures all components of emotion, and existing techniques that measure single components have limitations. Scientists have therefore suggested to use a combination of measurement techniques to cover more than one type of emotional response (Feldman Barrett, 2016; Kaneko et al., 2018). So far, verbal self-report has been used far more frequently than other, more implicit methods (Kaneko et al., 2018; Lagast et al., 2017). Due to the ease of application, ability to discriminate between products, as well as practical and cost effective aspects of self-report measurement, this will also be the method used to capture emotional responses in this thesis. It should therefore be noted that only conscious emotions will be measured and it is possible that responses will be subject to self-report bias and may be impacted by the act of measurement (Feldman Barrett, 2016; Kaneko et al., 2018). A reduced product and culture specific emotion lexicon with balanced positive and negative emotion categories will be used to capture emotional responses explicitly (Eaton et al., 2019). Using a reduced lexicon means that it will take a relatively short amount of time to complete the self-report task which is an advantage considering the fast nature of emotion (Coppin & Sander, 2016). Although the beer-specific lexicon has been validated, there is still a risk that certain emotional nuances will be missed due to the limited number of terms included (Eaton et al., 2019).

1.3 Context

1.3.1 Context and sensory consumer research: a validity trade-off

Despite the growing awareness that context is an important element to include in sensory consumer studies (Jaeger, Hort, et al., 2017; Köster, 2003; Meiselman, 2013), most research still takes place in controlled settings. This is partially due to a lack of research on the effects of how specific contextual factors impact test results (Galiñanes Plaza, Delarue, & Saulais, 2019), as well as concerns about losing experimental control (Meiselman, 2013). Indeed, taking research outside of controlled environments has consequences for the reliability and validity of the study. Reliability refers to the consistency of the test (Tavakol & Dennick, 2011). In other words, whether repeating the test would yield the same result (Lawless & Heymann, 1999). Validity is the ability of a test to measure what it was intended to measure (Lawless & Heymann, 1999; Tavakol & Dennick, 2011). Validity is assessed from two different perspectives: internal validity and external validity. Internal validity indicates the ability of experimental data to provide understanding and explain causal relationships within the experiment, whereas external validity is the ability of the test results to be generalized to situations outside of the test environment (Galiñanes Plaza et al., 2019). In contextual research there is sometimes referred to ‘ecological validity’. This term is not clearly defined (Jaeger & Porcherot, 2017; Schmuckler, 2001) but is commonly used to indicate the extent to which the environment in which a test takes place is representative of the context of interest (Galiñanes Plaza et al., 2019). Hence, a higher ecological validity is believed to increase external validity. Moving from controlled settings to testing in natural consumption environments is

considered a trade-off between internal and external validity (Galiñanes Plaza et al., 2019; Meiselman, 2013). Research usually relies on controlled environments to demonstrate causal relationships without any disturbance by external factors, in order to ensure reliability and internal validity. For this reason, it is common practice in sensory science to test products in isolated sensory booths in a laboratory or a central location test setting, minimising external sensory cues (Bangcuyo et al., 2015; Lawless & Heymann, 1999) to focus the individual's attention on the sensory characteristics of the product itself (Kemp et al., 2009; Lawless & Heymann, 1999). However, this is not a valid representation of any situation a consumer might normally experience when interacting with the product in 'the real world' (Bangcuyo et al., 2015; De Graaf, Cardello, et al., 2005). Hence, testing in controlled settings to optimise reliability and internal validity comes at the expense of external validity. The lack of ecological validity is cited as one of the reasons for high failure rates of new product launches, along with the lack of engagement of participants due to boredom in traditional central location tests (Bangcuyo et al., 2015).

1.3.2 Defining context

The relevance of context in food acceptance was discussed by Rozin and Tuorila (1993), who defined context as: *"a set of events and experiences that are not part of the reference event that one is interested in but have some relationship to it."* If these are physical factors happening at the same time as the reference event, they are referred to as simultaneous contextual factors, whereas memories of past events and anticipated future events are considered temporal contextual factors (Rozin & Tuorila, 1993). Meiselman (1996)

adapted Rozin and Tuorila's (1993) definition and defined the reference event to include the sensory, physiological and behavioural responses to the reference food, while the context includes those factors which surround the reference event and influence it. In this thesis the reference event is 'drinking beer', which according to Meiselman (1996) includes participants' emotional response to beer. As will be discussed in more detail in the next section, the contextual factors related to this reference event include the individual that is drinking beer, when, where, how and with whom the reference event takes place, as well as the properties of the beer itself and the way it is presented.

1.3.3 Organising contextual factors

1.3.3.1 Seven categories of contextual factors

Rozin and Tuorila (1993) emphasised the importance of contextual variables in determining food choice and liking of food and organised contextual effects as: 'simultaneous and temporal', 'food and non-food', 'single exposure (bite or sip) and a meal or pattern of eating'. Meiselman (1996) on the other hand provided a comprehensive review of context literature related to food acceptance and choice, and organised context factors as related to the 'Food', the 'Situation' and the 'Individual'. In a more recent overview Köster (2009) discussed the many factors that influence food choice from a psychological perspective and summarised the central question in food choice research as: ''Why does who eat what, when, and where?". This section will build on from these previous organisations of contextual research by proposing a structure that organises contextual factors in seven categories, as summarised in Figure 1.2.

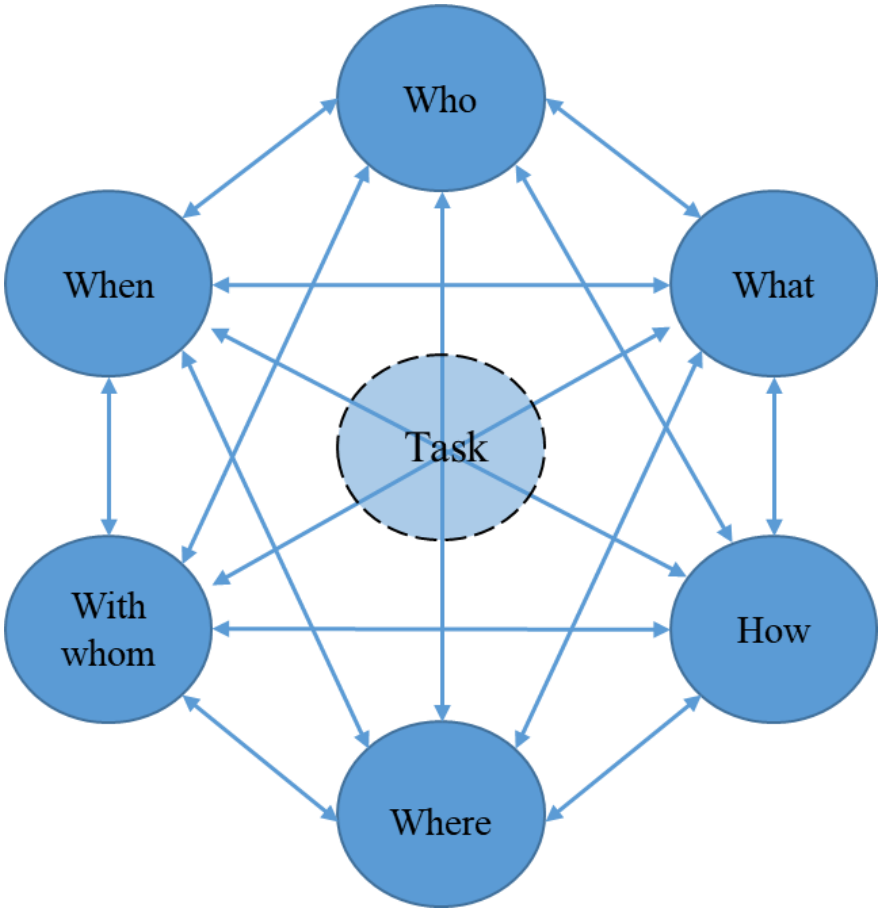


Figure 1.2: Overview of categories of contextual factors that impact consumers' responses to food. The individual, (Who) the food (What and How) and the situation (When, With whom and Where) are always present in consumption situations. The Task becomes relevant when the food consumption is part of sensory consumer research.

845

Six of the context categories represent factors that are always present when individuals interact with food, namely:

1) The individual (Who),
2) The stimulus (What),
3) The presentation, with or without additional information (How)
4) The moment in time (When)
5) The physical environment (Where)
6) Social interaction (With whom)

A seventh category, which is not applicable in everyday consumption but becomes very important in research situations, is the method used to collect data: the Task (Jaeger & Porcherot, 2017; Schmuckler, 2001). The organisation of contextual factors in Figure 1.2 can be used to systematically consider the contextual factors that are always present in sensory consumer research. A standardised sensory test setting is not void of contextual factors which is easily demonstrated with an example. When considering a CLT for beer, all seven contextual factors are relevant. A sensory researcher will need to decide on the serving temperature of the beer samples (What) and whether to present small amounts in plastic cups or glass jars, or larger amounts in glassware (How). The environment of a sensory lab differs from typical real life beer consumption environments (Where) and is likely to be unfamiliar to participants that have never participated in a sensory test before. How do the participants usually drink beer, and what are their expectations (Who)? Will participants like the beer more if they are feeling relaxed (Who) and would they have a more positive response to the beer in on a Friday than a Monday (When)? How will the participant respond to being isolated in a sensory booth,

and will their responses be impacted by the other participants or the person serving their beer (With whom)? Will the participant be required to rate all 39 emotions of the EsSense profile, or remember to look into a camera so that their facial expressions can be analysed (Task)?

Similar to Meiselman (1996) the proposed organisation includes the Food (What & How), the Situation (Where, When and With whom) and the Individual (Who) as contextual factors that could provide information about Why the reference event occurs. Adding the Task and separating Food and Situation into smaller categories provides a more detailed overview of contextual factors that are relevant to sensory consumer studies. Since context impacts liking and food choice, it can be hypothesised that context also impacts emotional response. In the next section each category of contextual factors will be discussed. The aim is to provide an overview of the impact of contextual factors in each category on consumers' liking, choice and emotional response to food and drink products. In addition, the areas where important information is lacking will be highlighted.

1.3.3.2 Interactions between different categories of contextual factors

In the next section different contextual categories will be discussed in isolation but in reality they impact each other, which is represented by the arrows in Figure 1.2 that connect each category of contextual factors to the others. The categories of contextual factors tend to overlap and are heavily related to each other, and so it is difficult to study their effects independently. The impact that their combined exposure has can be related to appropriateness (Jaeger & Porcherot, 2017; Piqueras-Fiszman & Jaeger, 2015). For example, whether the consumption situation (Where and When) is considered appropriate will

depend on the product (What and How), the individual (Who) and the social context (With whom). Schmuckler (2001) commented that *"the relevance of the environment is also critical in that it must be functionally central in producing and observing the behavior in question; it must be a true environment in which actors behave on a regular basis."* Thereby indicating the clear dependency between environment and individual.

1.3.3.3 Who

The importance of considering the variation in responses across individuals has been previously discussed by Köster (2003), who criticised the assumption that all participants are comparable, calling it the "fallacy of consumer uniformity". An increasing number of studies have focussed on investigating individual variation (Jaeger, Hort, et al., 2017). The individual's reactions are part of the reference event, whether these are sensory, emotional, psychological or physiological (Meiselman, 1996; Rozin & Tuorila, 1993). Attitudes, expectations and cognitions that the individual has prior to the occurrence of the reference event are part of the context (Meiselman, 1996). Attitudes, expectations and cognitions in their turn are influenced by many contextual factors such as the individual's genetic make-up, personality traits, mood, expectations and cultural background. Individuals differ in taste sensitivity based on their genetics, such as the ability to taste the bitterness of PROP (6-n-propylthiouracil) which has been linked to higher sensitivity to bitterness and other tastes (Yang, Dorado, Chaya, & Hort, 2018). Another example is thermal taste, which relates to the perception of tastes when the tongue is rapidly warmed or cooled. Both PROP taster status and thermal taster status have been shown to impact emotional response to beer, and PROP taster status also

impacted beer liking (Yang et al., 2018). Furthermore, there is a whole range of personality traits such as sensitivity to reward, to punishment and to disgust, private body consciousness, alexithymia and food neophobia that can be measured and linked to food preferences (Spinelli et al., 2018). Alexithymia, which relates to difficulty with describing feelings might impact the ability of participants to report emotional response. Neophobia and neophilia relate to an individual's willingness to try novel foods. Since novelty is one of the criteria involved with appraisal mechanisms (Coppin & Sander, 2016) emotional response is likely to differ between high and low neophobic participants. Variety seeking, which is an individual's sensitivity to monotony and the desire for variety, has also been related to food choice behaviour (Meiselman, 1996; Rozin & Tuorila, 1993). It is not known if variety seeking plays a role in consumption behaviour related to beer, but if it does, this tendency is likely to vary between individuals. Mood and emotional state can also cause variation in responses between individual participants, since they impact motivation to eat, food choice and affective responses to foods (Köster & Mojet, 2015; Macht, 2008). Macht (2008) reviewed literature on how emotions affect eating and emphasised that there is variability between individuals in how emotional states impact their food-related behaviours. In a sensory consumer study on wine, Danner et al. (2016) found that mood only had a modest effect on liking, but correlated moderately with emotional response, indicating that mood impacts emotional response. Changes in mood can be measured in a natural consumption context and appear to be product dependent (Porcherot, Petit, Giboreau, Gaudreau, & Cayeux, 2015). Participants' mood state before stimulus exposure is not commonly measured or taken into account in sensory

consumer research. The same is true for physiological states such as hunger, thirst or fatigue which are likely to impact perception and expectations (Köster, 2009), as well as emotional responses (Jiang et al., 2014). Previous experiences with a product or product category can influence expectations, which in turn can affect perceptions and liking (Rozin & Tuorila, 1993). Consumers' expectations can lead to assimilation-contrast effects, either minimising (assimilation) or magnifying (contrast) differences between product performance and expectation (Deliza & MacFie, 1996). Cultural background also plays a role in forming expectations and is known to affect emotional associations to products (Van Zyl & Meiselman, 2015, 2016) as well as consumption situations associated with beer (Silva et al., 2015). Cultural context is therefore a relevant factor in emotion research. With all the internal factors that impact an individual's response, variation between participants and within the same participant on different occasions is to be expected when measuring emotional responses.

1.3.3.4 What

Rozin and Tuorila (1993) mentioned several contextual factors related to the food itself such as the colour, sweetness and flavour of beverages and dynamic contrast in sensory characteristics that change during consumption. These are intrinsic attributes, which are the product's sensory, chemical and physical properties that cannot be changed or experimentally manipulated without also changing the physical characteristics of the product itself (Asioli et al., 2017). For beer, the colour can impact the expectations of beer consumers (Carvalho, Moors, Wagemans, & Spence, 2017) and serving temperature has been shown to impact emotional response (Dorado, Chaya, et al., 2016). Furthermore,

intrinsic properties for beer such as temperature and carbonation can interact as shown by Yang et al. (2018); at a standard carbonation level, cold beer was liked more and associated with a more positive emotional response than beer served at room temperature. For beer with a lower level of carbonation temperature had no effect. Piqueras-Fiszman and Jaeger (2014a) found that responses to more emotion-laden product categories were more variable between consumption situations than more neutral product categories. Consumers have been shown to hold strong emotional associations for beer (Chaya et al., 2015; Eaton et al., 2019; Jaeger, Cardello, et al., 2017; Silva et al., 2016) and thus this product category is likely to yield varying emotional responses in different consumption environments.

1.3.3.5 How

The way a product is presented can impact consumer responses. Product presentation entails meal context, tableware, extrinsic product attributes and any information about the product. The context of a meal can affect acceptability and preference for food (Rozin & Tuorila, 1993). Some food and drink items yield higher liking scores when presented as part of a meal than when served as individual items (King, Meiselman, Hottenstein, Work, & Cronk, 2007; King, Weber, Meiselman, & Lv, 2004). Within sensory test situations, the context of the sample within the sample set will impact perception due to contrast and adaptation effects (Lawless & Heymann, 1999) and the order in which samples are presented has been shown to have a strong effect on self-reported emotional response to beer, such that a sample presented in the first position tends to receive higher scores on emotions that are associated with positive valance, and lower scores on emotions associated

with negative valance (Dorado, Pérez-Hugalde, et al., 2016). Several studies have demonstrated the effect of tableware on liking and sensory perception of food (Piqueras-Fiszman, Alcaide, Roura, & Spence, 2012) and drinks (Spence & Van Doorn, 2017). For example, the type of glassware can influence participants' liking and willingness to pay for beer and other alcoholic drinks (Wan, Zhou, Woods, & Spence, 2015). Extrinsic attributes are aspects such as brand, price, packaging and health claims about the product which are not a part of the physical product itself (Asioli et al., 2017). Labelling can affect both liking and emotional response (Ng, Chaya, & Hort, 2013b; Silva et al., 2017). Even label information that is known to be false can have a significant impact on preference and choice behaviour (Rozin, Markwith, & Ross, 1990). The effect of labelling on expectations and assimilation/contrast effects in terms of liking and emotional response has been demonstrated with beer products (Silva et al., 2017). Moreover, the strong effect of branding on consumer evaluation for beer has been known for a long time (Allison & Uhl, 1964). Furthermore, information such as price and beer style has been shown to influence consumers' perceptions of beer quality (Aquilani, Laureti, Poponi, & Secondi, 2015). A food can evoke feelings of disgust simply due to the knowledge that something undesirable has happened to it prior, even though the food is otherwise edible according to current sensory information (Rozin & Tuorila, 1993). This demonstrates how the way products are presented, and with what kind of information, clearly impacts consumers' responses.

1.3.3.6 Where

The vast majority of context research has focussed on this category of contextual factors that represents the physical environment in which sensory

consumer research can take place. Sensory consumer research can be conducted as a Central Location Test (CLT), Home Use Test (HUT), or in a natural consumption environment. The CLT is the most common practice and involves inviting participants into a controlled environment, either a sensory laboratory or other central location, for product testing (Lawless & Heymann, 1999). With the HUT, participants are provided with the products at home to use repeatedly under common daily circumstances (Boutrolle, Delarue, Arranz, Rogeaux, & Koster, 2007; Jaeger & Porcherot, 2017). Natural consumption environments outside of the home commonly used for sensory consumer testing include restaurants (Edwards, Meiselman, Edwards, & Lesher, 2003; Porcherot et al., 2015), bars (Gomez-Corona, Chollet, Escalona-Buendia, & Valentin, 2017; Silva et al., 2017) or cafeterias (Edwards et al., 2013; Hedderley & Meiselman, 1995; Petit & Sieffermann, 2007). CLT is less costly and easier to organise than HUT and testing in natural consumption environments, hence its frequent use with food products (Jaeger & Porcherot, 2017). CTL results generally underestimate consumer's acceptability of products in real life environments (Jaeger & Porcherot, 2017; King et al., 2007; King et al., 2004; Meiselman, Johnson, Reeve, & Crouch, 2000), although this effect varies between products (Boutrolle & Delarue, 2010; Boutrolle et al., 2007). As mentioned in section 1.3.1, choosing between controlled environment and real life situations for sensory consumer testing is a trade-off between internal and external validity. A realisation of the importance of context has led to the use of different ways of bringing relevant context to controlled test settings as an attempt to optimise both experimental control and ecological validity. Different strategies include written scenarios

(Dorado, Chaya, et al., 2016; Hein, Hamid, Jaeger, & Delahunty, 2010), exposure to images (Andersen, Kraus, Ritz, & Bredie, 2019; Hersleth, Monteleone, Segtnan, & Naes, 2015), video or sound recordings (Hathaway & Simons, 2017; Liu, Hannum, & Simons, 2019), and decorating rooms to be more similar to real life environments (Holthuysen, Vrijhof, De Wijk, & Kremer, 2017), sometimes with video projection (Sester et al., 2013; Sinesio et al., 2018). Technological developments have led to the use of video walls (Bangcuyo et al., 2015; Hathaway & Simons, 2017), and virtual reality (Andersen et al., 2019; Sinesio et al., 2019). Most sensory consumer research on context has looked into effects of environment on liking (Boutrolle & Delarue, 2010). The test environment appears to have varying effects, depending on the product under testing (Boutrolle & Delarue, 2010; De Graaf, Cardello, et al., 2005). For beer, situational appropriateness appears to be related to familiarity with the product (Giacalone et al., 2015), indicating that novelty might play a role in consumers' responses to beer in different situation. It has already been demonstrated that changing the atmosphere of an environment can be used to manipulate the choice for beer and other beverages (Sester et al., 2013). Furthermore, product differentiation on emotional response to beer can be increased by evoking a relevant consumption context in a controlled setting with a personalised written scenario (Dorado, Chaya, et al., 2016). Sinesio et al. (2019) found that product differentiation on liking and emotional response to beer products was stronger in a CLT setting than in a real pub and when simulating contexts with an immersive room and virtual reality. Further research is needed to investigate the stability of emotional responses to beer across different test environments.

1.3.3.7 With whom

The presence of other people has been shown to impact eating behaviour. For example, the time taken to eat lunch increased with the number of people at the table (Bell & Pliner, 2003) and eating with other people can increase the size of a meal (Patel & Schlundt, 2001). Exposure to facial expressions of other people has been shown to impact the desire to eat foods (Barthomeuf, Rousset, & Droit-Volet, 2009). Nonetheless, social interaction does not appear to have a strong effect on liking for a range of foods. For example, hedonic scores were not observed to be impacted when participants interacted with others, during the consumption of iced tea, salad, breadsticks, cannelloni, lasagne (King et al., 2007; Weber, King, & Meiselman, 2004) or cheese (Hersleth, Ueland, Allain, & Naes, 2005). However, as an alcoholic beverage, beer consumption is associated with social situations (Keesman et al., 2018) and thus, social interaction could be a relevant factor for emotional response to beer.

1.3.3.8 When

The time of day can have a strong impact on food preference, depending on the perceived appropriateness (Birch, Billman, & Richards, 1984). Consumers will respond to a brownie or an apple differently depending on whether they imagine having it for a ‘weekend breakfast’, an ‘afternoon snack’ or ‘after a special dinner’ (Piqueras-Fiszman & Jaeger, 2014b). King, Meiselman, and Carr (2013) found no effect of the time of day on the emotional response for salty snacks, but noted that this is a product that is usually eaten at any time of the day. For beer, the emotional response is more likely to vary between different times a day, days of the week and even between different times of the year, due to situational appropriateness. The time of day had an effect for fruit-

flavoured non-alcoholic beers in an experiment by Delarue, Brasset, Jarrot, and Abiven (2019) who found higher hedonic ratings during afternoon sessions than morning sessions in the sensory lab. Logically, presenting a beer at an appropriate time, such as 8pm on a Friday, is likely to yield a higher liking score and more positive emotional response than when the same product is presented on a Monday morning. Therefore, this contextual factor should be taken into account when emotional responses to beer are measured.

1.3.3.9 The Task

As mentioned in section 1.2, emotion measurement can change emotion, and therefore the task should be taken into account as a contextual factor in sensory consumer research. In most sensory consumer test participants are not free to choose the products they wish to evaluate, even though this has demonstrated consequences for acceptability of some food products (De Graaf, Cardello, et al., 2005; King et al., 2007; King et al., 2004). Additionally, the awareness of being part of an experiment could impact participants' behaviour, hence the advantage of implicit methods. King et al. (2013) studied the effect of specific test and questionnaire design elements on self-reported emotional response. They concluded that including more than two samples resulted in responses that were more associated with the task than the samples. In addition, rating scales were found to be more sensitive in detecting product differences than check-all-that-apply (CATA), although CATA was faster and easier for participants to use.

1.4 The impact of context on emotional response

Emotion and context have received increasing interest from the sensory field (Jaeger, Hort, et al., 2017; Meiselman, 2013). Based on the emotion theories discussed in section 1.2 emotions are context-dependent. Nonetheless, most studies that measured emotion so far have taken place in a central location such as a sensory laboratory (Lagast et al., 2017). Both emotional response and context appropriateness seem to be product specific and thus there is a need to research these phenomena within specific product categories. Beer is associated specific consumption situations (Giacalone et al., 2015; Silva et al., 2015) and has clear emotional associations (Chaya et al., 2015; Eaton et al., 2019; Silva et al., 2016; Silva et al., 2017). Therefore, the emotional responses consumers experience to beer is expected to vary between contexts. In addition, variation is expected across individuals and within the same individual over time, based on emotion theory (Achar et al., 2016; Coppin & Sander, 2016; Feldman Barrett, 2016) and the many factors that determine individual variation discussed in section 1.3.3.3.

1.5 Overview of thesis content

This introduction chapter clarified the different elements that make up emotion and context. The rest of this thesis comprises four studies that will each be presented in a different chapter, followed by an overall conclusion in the final chapter. The study in Chapter 2 was published in Food Quality and Preference in January 2019 and will therefore be referenced as Nijman, James et al (2019) in other parts of this thesis. Other result chapters have been written as paper manuscripts but have yet to be published, and so where reference to results presented in these chapters is made in the thesis, they are referenced as Chapter 3, 4 or 5.

Chapter 2 describes a study in which consumer responses to two beers obtained in a bar, and when evoking a bar context in a sensory laboratory were compared to responses from a standard sensory laboratory setting. This study aimed to determine the effect of each context condition on product differentiation in terms of emotional response, liking and product choice, and on the variation in context sensitivity amongst beer consumers.

A second study, which aimed to discern the changes in emotional response and liking for beer over different consumption stages is written up as a short communication paper in Chapter 3. The study determined the consequences of a single-sip approach versus the consumption of a representative quantity of beer. In addition, it is unknown if emotional responses to products, especially beer, which elicits a strong emotional response, are consistent and repeatable. Therefore, a secondary aim of this study was to explore the reliability of emotional responses to beer over repeated sessions.

A small-scale study, which is outlined as a short communication paper in Chapter 4, addressed concerns related to the effect of using different devices to collect emotional response data in the bar and controlled setting during the first two experiments. Emotional responses and liking were measured with the same questionnaire, presented on two different devices; a desktop computer in the sensory booth, and the participants' personal smartphone.

Chapter 5 describes the final study, which explored the use of Virtual Reality (VR) to simulate a bar context in a controlled environment. Qualitative research methods were used to investigate participant engagement and practical considerations involved with VR as a novel technology. Furthermore, participant engagement and emotional responses to two beers were compared between VR and evoking the same bar context in sensory booths with pictures and sound recordings.

In Chapter 6 conclusions are drawn based on the findings from all experiments and recommendations are made for future research.

Chapter 2

2. The effect of consumption context on consumer hedonics, emotional response and beer choice

Preliminary thoughts Chapter 2:

"According to the present authors, the segmentation of the consumers on the basis of their preference is a major step forward and should be recommended in all food-related emotion research"

Köster and Mojet (2015)

Based on emotion theory and the impact of contextual factors on consumers' responses discussed in the introduction chapter, emotional responses to beer were expected to vary across contexts, as well as across individuals. The experiment detailed in this chapter explored variation across participants in self-reported emotional responses to beer, measured in a sensory laboratory, a real life consumption environment and when evoking a relevant consumption context in a sensory laboratory.

This chapter was published as a paper in Food Quality and Preference in January 2019:

Nijman, M., S. James, F. Dehrmann, K. Smart, R. Ford and J. Hort (2019). "The effect of consumption context on consumer hedonics, emotional response and beer choice." Food Quality and Preference 74: 59-71.

The effect of consumption context on consumer hedonics, emotional response and beer choice

Marit Nijman[1], Sue James[2], Frieda Dehrmann[2], Katherine Smart[2], Rebecca Ford[1], Joanne Hort[1, 3]

[1] *Division of Food Sciences, University of Nottingham, Sutton Bonington Campus, United Kingdom*
[2] *Anheuser-Busch InBev*
[3] *Riddet Institute, Massey University, Palmerston North, New Zealand*

Key words: Context, Emotional response, Liking, Product Choice, Beer

Highlights:

- Consumer response to beer was affected by context.
- Some consumers maintained their response to beer independent of context.
- Some consumers were more context sensitive when expressing affective response.
- Context sensitive consumers discriminated products in the bar but not in the lab.
- Emotional variables provided additional insights compared to liking.

2.0 Abstract

The context of a consumer test affects participant response. Data collected in a sensory laboratory is likely to have little predictive value of consumer experience in real life situations. This study determined the effects of context on consumer response to two commercial beers. Regular beer consumers (N = 100) rated liking and emotional response using ten beer-specific emotion categories for two beers (Lager and Ale) under three different conditions: (1) a sensory testing facility (Lab), (2) a natural consumption environment (Bar) and (3) using an evoked context (Evoked). Their choice of product to take home was also recorded. Overall results showed significant product differentiation for liking (F (99, 2, 1) = 8.46, p = 0.004) and product choice (Q (1, N = 100) = 4.85, p = 0.028) in the Bar but not in the Lab or Evoked context. Emotional variables highlighted significant product differentiation ($p < 0.05$) but more so in the Bar than in the Lab or Evoked context. However, clustering participants on liking revealed three distinct clusters differing in sensitivity to context. Two clusters showed opposing but consistent preference for one of the two products regardless of context. The third cluster was more influenced by context, showing a more discriminating response in the Bar. These findings showed that consumers differ in their degree of context-sensitivity and the extent to which evoking a context gives similar results to a real environment. They also highlighted the importance of segmentation and confirmed the added insights gained by measuring emotional response compared to liking.

2.1. Introduction

Not taking the situation in which a product is consumed into account has been mentioned as a common fallacy in sensory science and consumer research (Köster, 2003). Eating is a multisensory experience (Spence, 2013) and human beings are influenced by the context and environment in which they consume products (Bell, Meiselman, Pierson, & Reeve, 1994; Edwards, Hartwell, & Giboreau, 2016). Good practice in sensory research dictates that a lot of effort goes into minimising variation due to external sensory signals by having white/neutral walls and furniture, an odourless environment and temperature and humidity control to avoid introduction of experimental error. While this setting might be preferred for objective testing, it provides an unrepresentative consumption situation for consumers and therefore is likely to have little predictive power for how consumers experience products 'in the real world' (De Graaf, Cardello, et al., 2005). It has also been argued that testing in isolated sensory booths leads to boredom and a lack of attention amongst participants, thus further diminishing the external validity of consumer data (Bangcuyo et al., 2015). Since the setting in which a sensory consumer test is performed can affect the way participants respond to the tested products, providing participants with a context that is closer to the natural consumption situation of the tested products is an area of increasing interest for sensory researchers (Jaeger, Hort, et al., 2017). One strategy to improve validity of sensory consumer tests has focussed on the use of 'evoked contexts' by describing meal situations in a written scenario (Piqueras-Fiszman & Jaeger, 2014c), asking consumers to write their own scenario (Dorado, Chaya, et al., 2016; Hein et al., 2010) or by re-creating a natural consumption environment

by manipulating elements in a controlled setting (Bangcuyo, Smith, Zumach, Pierce, Guttman, & Simons, 2015; Holthuysen, Vrijhof, de Wijk, & Kremer, 2017; King, Weber, Meiselman & Lv, 2004).

Context is known to impact food choice and acceptability. Studies comparing hedonic ratings obtained in laboratory and real life consumption environments found that, depending on the product type, liking scores can vary between sensory laboratories and Home Use Tests (Boutrolle et al., 2007), restaurants (King et al., 2007; Meiselman et al., 2000) and cafeterias (Meiselman et al., 2000) respectively. It has also been demonstrated that consumer choice behaviour can be manipulated by the decoration in a restaurant (Bell et al., 1994) or the ambiance of a bar (Sester et al., 2013). Several authors suggest context affects emotional response as well (Edwards et al., 2016; Piqueras-Fiszman & Jaeger, 2014b), however, limited research has been done in this area. Although there is no consensus amongst scientists on the definition of emotion and this topic is subject to debate in the field of sensory and consumer science, the measurement of 'emotions' has gained popularity in recent years. Coppin and Sander (2016) discuss a consensus definition based on the current emotion literature: an emotion is an "event-focused, two-step, fast process consisting of (1) relevance-based emotion elicitation mechanisms that (2) shape a multiple emotional response (i.e., action tendency, automatic reaction, expression, and feeling)". This definition was adopted for the purpose of defining emotion in the current research, assuming that exposure to the beer causes a relevance-based emotion elicitation mechanism and that a self-report measure can capture the 'feeling' aspect of the multiple emotional response. It has been shown that measuring consumer emotional response using self-report

methods is more discriminating than simply measuring liking (Ng et al., 2013a) and may provide deeper insights into food choice decisions (Gutjar, Dalenberg, et al., 2015; Gutjar, De Graaf, et al., 2015). Therefore, there is a need to determine if and how context affects emotional response data in order to improve practice in consumer testing of products. Piqueras-Fiszman and Jaeger (2014b) found that asking consumers to imagine eating the same food in different contexts using written scenarios changed their emotional response depending on their perceived appropriateness of the evoked consumption occasion. Dorado, Chaya, et al. (2016) found that imagining a consumption context with the use of a written scenario impacted emotional response to beer compared to testing without evoking a context. However, Jiang, Niimi, Ristic, and Bastian (2016) studied the effects of decorations in an immersive environment on the emotional response to wine and found that the immersive environment had no effect. Despite contradicting evidence on the effects of context, some researchers choose 'real life' settings to measure liking and emotional response to beer (Gomez-Corona et al., 2017; Silva et al., 2017). Beer as a product category is associated with positive high arousal emotional responses (Silva et al., 2016) and the use of an emotion lexicon has been shown to discriminate between beer samples (Chaya et al., 2015). Hence, beer can be considered an effective stimulus to study product related emotional responses.

There have been no previous studies that compared emotional response to beer obtained in a laboratory setting to a natural consumption context or evoked context. The current study aimed to fill the gap of knowledge in this area by comparing consumer liking, emotional response and choice of beer products in

a natural consumption context, i.e. a bar, to a traditional sensory test setting, and to determine the relative effect of evoking a context. Regular beer consumers were invited to evaluate beer and complete a questionnaire in a bar, a sensory lab and under an evoked context condition in three separate sessions. The study was approved by the Faculty of Medicine & Health Sciences Research Ethics Committee of the University of Nottingham (Ethics reference number H12092016).

2.2. Materials and methods

2.2.1 Products

A commercial ale and lager beer were selected as sensory stimuli. These two products were selected for being visually similar but noticeably different in taste, flavour and mouthfeel. Similarity in appearance was deemed necessary to prevent differences in results due to visual cues. The products were characterised by The University of Nottingham trained beer panel as having different sensory profiles. Eight of these panellists generated sensory attributes to describe the beer and performed 2-Alernative Forced Choice (2-AFC) tests comparing the two products on each attribute for intensity in triplicate. The Ale was found to be significantly stronger in fruity aroma ($p = 0.003$), bitter taste ($p < 0.001$) viscosity (mouthfeel) ($p = 0.001$) and astringent mouthfeel ($p = 0.003$) than the Lager. The Lager was found to be significantly stronger in sulphury aroma ($p = 0.011$) and apple flavour ($p < 0.001$) than the Ale. The Ale had an alcohol content of 4.1% Alcohol by Volume (ABV). The Lager beer contained 4.8% ABV.

2.2.2 Participants

One hundred regular beer consumers (consuming beer at least once a month) were recruited from staff (15%), and student (82%) volunteers at the University of Nottingham, plus some locals external to the University (3%). The participants (40% female) ranged in age from 18 to 65 with an average age of 25 years. Pregnant women and individuals that had any reason to refrain from drinking alcoholic beverages (including declared health, religion or addiction) were excluded from participation. Seventy-four participants (74%) were from the UK and 87% declared that they spoke English as their first language. Non-UK participants were from Europe (11%), Asia (11%), the Americas (3%) and Africa (1%) and declared that they had been living in the UK for at least one year. Participants received an inconvenience allowance in recompense for their time.

2.2.3 Experimental design

Participants attended three sessions over the course of five weeks during which they were asked to taste two beer products and record their responses via an online questionnaire. To balance out session order effects over the three context conditions, participants were divided into three groups based on their availability resulting in 39 participants having their first session in the Bar, 33 in the Lab and 22 starting with the Evoked context. Because session order was partially based on participants' availability, the order of the contexts was not completely balanced. The number of participants per each of the six possible session orders for attending the Lab, Bar and Evoked contexts ranged from 10 to 22. All test sessions took place on Mondays, Tuesdays and Wednesdays in 30-minute timeslots between 5pm and 8:30pm. As much as individual

schedules allowed, all three sessions were scheduled for the same time and on the same day of the week for a specific individual participant.

2.2.4 Context conditions

Data was collected during three sessions, each under a different context condition. A Student Union bar at the University of Nottingham (Figure 2.1) was used as the natural consumption context (Bar) for one session. The usual ambiance of the bar was unchanged during the experiment and the bar remained open as usual for other customers. After receiving test instructions participants were free to sit anywhere in the bar. Talking was allowed during the test, but instructions were given to not discuss the questionnaire or the samples. In order to mimic a natural consumption context as much as possible, samples were prepared behind the bar and participants were instructed to pick up their beer from the bar at specified moments during the questionnaire. Questionnaires were completed on participant's smartphones. Tablets were available for participants that did not have a smartphone. Since it is common to use a smartphone in a bar this was considered to cause minimal disturbance to the natural setting. The other two sessions took place in the Sensory Science Centre sensory booths (ISO: 8589:1988) at the University of Nottingham (Figure 2.1), one under standard conditions (Lab) and another where the context was evoked (Evoked). Under the Lab and Evoked context conditions questionnaires were completed on desktop computers in the sensory booths. In the Evoked context condition participants were asked to imagine that they were in the Student Union bar context condition while they were physically in the same sensory booths as under the Lab condition. Before receiving their first sample the participants were exposed to a written instruction describing the



Figure 2.1: On-campus Student Union bar at the University of Nottingham (left) and sensory test facility at the University of Nottingham (right)

Student Union bar and asked to imagine what it would be like to be there. To help participants to imagine the Bar context, they were exposed to sound recordings and pictures from the Student Union Bar on a tablet in their tasting booth. Thirty minutes of sounds were recorded on a regular Tuesday evening in the Student Union Bar and consisted of music, indistinguishable conversation and background noises. Five photos were taken from different angles in the Bar during a regular Tuesday evening and were displayed as a repeating slideshow with a duration of five seconds per photo. The slideshow and sound recordings played for the entire duration of the test, which took on average 10 minutes per participant. Participants were instructed to keep their earphones in and listen to the sound recordings until they had completed the entire questionnaire. No data was collected on compliance with instructions or the time participants actually spent watching the slideshow.

2.2.5 Sample preparation and temperature

Sample preparation initiated when participants started their questionnaire. Both beers were presented in identical standard half pint glasses (284 ml). For each sample a full glass containing approximately 284 ml was presented on a plain cardboard beer coaster labelled with a random 3-digit code. The temperature of the beer samples was recorded. As a result of limited cooling facilities and lack of control over the room temperature the average temperature of the samples in the Bar was relatively high at an average of 11.27°C (SD = 2.72). To avoid differences between the two locations, the sample temperatures in the sensory test facility were adjusted to match those in the Bar by storing the samples at room temperature for 30 minutes before each session. The average temperature of samples served in the sensory test facility was 10.17°C (SD = 1.56). The

average temperature of the Ale was 10.38°C (SD = 2.15) and for the Lager the average temperature was 10.87°C (SD = 2.11). To check for any significant differences in serving temperature between the two locations, and the two products, a two-factor Analysis of Variance (ANOVA) was performed (location, product) with interaction on the recorded sample temperatures. In the Bar the average sample temperature was found to be significantly higher than in the Sensory test facility (F (1, 1) = 6.26, p = 0.014). However, this significant difference was only due to the exceptional high average temperature that was recorded during the first test week in the Bar context (M = 12.66, SD = 2.16). There was no significant difference in temperature between the two products (F (1, 1) = 0.86, p = 0.367). ANOVA showed no significant interaction effect between context and product (F (1, 1) = 0.577, p = 0.449). It is unknown if the magnitude of the temperature differences in the beer samples would have been perceived by the participants. The temperature variations can be considered part of the experimental set-up, as in sensory test facilities there is more control over serving temperatures and the room temperature whereas in real consumption environments that level of control does not exist. Since the difference in temperature was caused by the first week of sessions, additional analyses were performed to check for session order effects.

2.2.6 Sample presentation

Beers were presented to participants monadically under blind conditions and in the same way across the three context conditions according to a randomised balanced design. A two-minute break was enforced between samples during which participants were instructed to cleanse their palate with mineral water.

Palate cleansing is not normally carried out in real life consumption situations but as two different beers were tasted close together in all contexts, palate cleansing was necessary to avoid carry over. It is important to note that participants were only allowed to take one sip of each sample. For this first investigation, one sip was deemed sufficient to measure participants' responses to the beers without the additional effects related to satiety and alcohol consumption. Future work will aim to determine the effects of consumed beer volume on consumer responses. The sip size was not controlled or standardised between participants as in this study the objective was to keep the consumption as natural as possible. Participants were instructed to take a sip as they would normally and thus the sip-size likely varied between participants.

2.2.7 Questionnaires

Data was collected using an online questionnaire via Compusense Cloud (Version 8.0.6288.23054, Compusense Inc., Guelph, Ontario, Canada). Participant's emotional response to each product was measured directly after tasting the beer using a previously developed beer specific emotion lexicon (Dorado, Chaya, et al., 2016; Eaton, Chaya, Smart, & Hort, 2018) consisting of ten emotion categories (Table 2.1). Each of the ten emotion categories was presented together with the associated terms and participants were asked to indicate the extent to which they were experiencing the emotions associated with those descriptors by giving an intensity rating on a continuous line scale anchored from 'very low' to 'very high' at 5% and 95% of the scale. The order of the emotion categories was randomised between different participants, but the order was kept the same over the three sessions for each participant. After evaluating emotional response, participants were asked to rate overall liking on

a continuous line scale, anchored at 5% and 95% with 'dislike extremely' and 'like extremely'. At the end of each test session participants were also asked which of the two tasted beers they would like to receive at the end of the study as a further thank you. They could indicate their choice by selecting one of the 3-digit codes of the samples tasted during that test session, or choose to express no preference for either product. At the end of their last session, participants received the bottle of the beer that they chose during that session. Although the choice was recorded during each of the three sessions, the participants only received the beer of choice after completion of the study, since presenting the beer would reveal the brand and product information which could have biased their responses in subsequent sessions.

Table 2.1: *Emotion categories and associated terms adapted from Dorado et al. (2016)*

Shocked	Shocked, alarmed, cheated, confused, overwhelmed, strange, weird
Bored	Bored
Content	Content, calm, comfortable, comforted, enjoyment, good, happy, nice, pleasant, pleased, relaxed, satisfied
Excited	Excited, enthusiastic, fulfilled, fun, impressed, interested, optimistic, pleasantly surprised, want, warm
Nostalgic	Nostalgic, delirious, relieved
Disconfirmed	Disappointed, dissatisfied, unpleasantly surprised
Disgusted	Disgusted, horrible, repulsed, repelled, unpleasant
Tame/safe	Tame, safe
Underwhelmed	Underwhelmed
Curious	Curious

2.2.8 Data analysis

Statistical analyses were performed using Microsoft Excel 2013/XLSTAT (XLSTAT Version 18.07.39020, Addinsoft, New York, USA). An α-risk of 0.05 was set as the level of significance in all data analyses, unless indicated otherwise.

Mean intensity scores and standard errors (SE) for all emotional categories and liking were calculated for each product (Ale and Lager) under each context condition (Bar, Lab and Evoked). To determine the overall effects of product and context, four-factor mixed model Analysis of Variance (ANOVA) with product, context and session number as fixed factors, and participants as a random factor, was performed for each emotion category and liking. Where significant product and/or context effects were found, Tukey post hoc analyses were applied. Session number (whether it was the first, second or third session for the participant) was included in the ANOVA model to account for the effect this might have had. As an aim of this study was to determine the consequences of decisions made regarding the context for consumer test designs, and given the lack of context*product*session interactions in the full data set, each context was also considered separately to demonstrate the impact of using a different context on product differentiation. Within each of the three context conditions, the data was split by context and a mixed model ANOVA with participant as random factor and product as fixed factor was performed on liking and emotional response from the Bar, Lab and Evoked context separately. To study differences in product choice per context Cochran's Q tests were performed comparing the number of times the Ale and the Lager were selected under each context condition. Participants not expressing a preference were excluded from the analyses.

A hierarchical agglomerative cluster analysis using Wards method was performed using product liking scores across all contexts to determine if liking patterns were homogeneous amongst participants. The resulting dendogram was visually inspected to determine obvious clusters present. The data were

then split by cluster. Mean scores and standard errors (SE) were calculated for each of the ten emotion categories and liking for each product under each context condition separately. To determine product and context effects within each cluster, a four-factor mixed model ANOVA with product, context and session number as fixed factors and participants as random factor was performed for each cluster separately. To determine how context affected liking and emotional response per product within each cluster, the data was split per product and analysed with a three-factor mixed model ANOVA with participant as random factor and context and session as fixed factors, separately for each cluster. The choice behaviour of each cluster was analysed per context with Cochran's Q test, as above. The emotion data, separated by product, context and cluster was subjected to a principal component analysis (PCA) in order to enable visualisation of the product-context-cluster emotional space.

2.3. Results

Emotional response, liking scores and preferred product choice for two commercial beer products were collected from 100 consumers in three different context conditions. Here, the first section describes relevant findings from analysis of the complete dataset (all 100 participants), the second section presents separate analyses of three distinct consumer clusters discovered in the data.

2.3.1 Overall consumer response to beer in different contexts

2.3.1.1 Context, product and interaction effects when considering all participants

Table 2.2 displays the mean scores per product for liking and each of the ten emotion categories when evaluated by 100 participants in the Bar, Lab and Evoked contexts. Table 2.3 presents the *F* and *p*-values from the four-factor ANOVAs for each emotion category and liking with all 100 participants were considered. As is expected in consumer studies, significant differences amongst participants were evident for all emotion categories as well as liking. The analysis over all participants revealed limited effects of context on consumers' responses to the two beers. A significant context effect ($p < 0.05$) was found for emotion category Curious, and context effects for Tame/safe and Underwhelmed were approaching significance ($p < 0.10$). As can be seen in Table 2.2, average scores on Curious were higher in the Bar than in the other two context conditions and Tukey post hoc groupings confirmed that the difference between the Bar and Lab was significant ($p = 0.006$) while the difference between the Bar and Evoked context approached significance ($p = 0.070$) for this emotion. The ANOVA results in Table 2.3 also showed a significant product effect ($p < 0.05$) for overall liking, as well as for emotion categories Shocked, Content, Nostalgic, Disgusted and Tame/safe. As can be seen from the means in Table 2.2, the lager was scored higher on liking and the significant positive emotions, whereas the Ale received higher scores for the significant negative emotion categories. Significant session effects were found for Shocked, Bored, Content, Excited, Disconfirmed, Disgusted, Underwhelmed, Curious and liking (Table 2.3). Tukey post hoc groupings

revealed that this was due to first order session effects, where participants gave a more positive response during the first session compared to the second and third sessions. No significant interactions were found between context and product or product and session (Table 2.3). The only significant interaction between context and session was found for Bored (Table 2.3). In the Bar slightly higher scores for Bored were given during the second session than during the first and last session. The cause of this effect is unknown. The only significant interaction between context, product and session was for the emotion category Disgusted (Table 2.3). The Ale was scored higher on Disgusted during the third session in the Bar context than during other sessions. It is not clear what caused this higher score.

Table 2.2: *Mean scores and standard errors (SE) for the Lager and Ale when evaluated by 100 participants in the Bar, Lab and Evoked context.*

		Bar				Lab				Evoked			
		Lager		Ale		Lager		Ale		Lager		Ale	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Emotional response	Shocked	1.73	0.15	2.58	0.24	1.57	0.17	2.25	0.25	2.09	0.22	2.48	0.26
	Bored	2.96	0.21	2.76	0.21	3.12	0.24	2.84	0.22	3.08	0.24	2.71	0.22
	Content	6.36	0.19	5.76	0.24	6.07	0.20	5.72	0.24	6.00	0.22	5.53	0.26
	Excited	5.46	0.21	5.13	0.26	4.98	0.23	5.05	0.25	5.02	0.24	4.96	0.23
	Nostalgic	4.51	0.24	3.87	0.26	4.26	0.26	3.87	0.26	4.10	0.26	3.86	0.25
	Disconfirmed	2.75	0.24	2.88	0.26	2.72	0.23	2.96	0.27	2.74	0.24	2.95	0.26
	Disgusted	2.06	0.19	2.87	0.26	2.04	0.21	2.43	0.26	1.87	0.19	2.51	0.26
	Tame/safe	6.28	0.21	5.41	0.26	5.96	0.23	5.09	0.25	5.55	0.26	5.25	0.27
	Underwhelmed	3.18	0.24	3.57	0.27	3.65	0.27	3.13	0.24	3.14	0.27	2.80	0.24
	Curious	4.90	0.24	4.76	0.25	3.95	0.25	4.44	0.26	4.16	0.26	4.59	0.25
	Liking	5.94	0.21	4.95	0.26	5.58	0.23	5.32	0.27	5.48	0.23	5.11	0.27

***Table 2.3**: F and p-values from a mixed model four-factor ANOVA (participant, context, product, session). Bold font highlights p-values < 0.05.*

		Main effects 4-factor ANOVA								Interaction effects 4-factor ANOVA							
		Participant (DF=99)		*Context (DF=2)*		*Product (DF=1)*		*Session (DF=2)*		*Context* Product*		*Context* Session*		*Product* Session*		*Context* Product* Session*	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	3.26	< **0.001**	2.23	0.109	17.85	< **0.001**	4.115	**0.017**	0.821	0.441	0.950	0.435	0.101	0.904	1.388	0.237
	Bored	3.55	< **0.001**	0.24	0.787	3.45	0.064	7.441	**0.001**	0.076	0.927	3.082	**0.016**	0.165	0.848	0.101	0.982
	Content	3.42	< **0.001**	0.76	0.471	9.13	**0.003**	4.665	**0.010**	0.269	0.764	0.971	0.423	0.425	0.654	1.173	0.322
	Excited	3.45	< **0.001**	0.92	0.398	0.40	0.526	4.066	**0.018**	0.512	0.599	0.595	0.667	0.009	0.991	2.189	0.069
	Nostalgic	4.67	< **0.001**	0.63	0.532	6.63	**0.010**	0.305	0.737	0.428	0.652	1.829	0.122	0.249	0.780	1.242	0.292
	Disconfirmed	2.12	< **0.001**	0.02	0.982	1.08	0.299	4.890	**0.008**	0.084	0.919	1.815	0.125	0.708	0.493	0.912	0.457
	Disgusted	2.46	< **0.001**	1.58	0.208	13.72	< **0.001**	6.376	**0.002**	0.520	0.595	0.485	0.747	0.523	0.593	2.673	**0.031**
	Tame/safe	5.84	< **0.001**	2.65	0.072	21.18	< **0.001**	1.420	0.243	1.703	0.183	0.819	0.513	0.165	0.848	0.807	0.521
	Underwhelmed	2.96	< **0.001**	2.77	0.063	0.75	0.386	3.421	**0.033**	2.218	0.110	1.231	0.297	0.226	0.798	0.794	0.530
	Curious	3.68	< **0.001**	4.29	**0.014**	2.40	0.122	4.375	**0.013**	1.437	0.239	1.354	0.249	0.010	0.990	0.286	0.887
	Liking	2.05	< **0.001**	0.17	0.844	8.65	**0.003**	4.275	**0.014**	1.458	0.234	1.446	0.218	0.047	0.954	1.891	0.111

2.3.1.2 Product differentiation per context when considering all participants

Figure 2.2 shows the average emotional response and liking for each product, separately in each of the three context conditions when all 100 participants were considered. Although little effect of context or context*product interaction was found when analysing the data of all context conditions together, there was a stronger product differentiation in the Bar context than in the Lab and Evoked context. When the products were tasted in the Bar, the Ale was scored significantly higher than the Lager on two negative emotion categories (Shocked and Disgusted) and lower on liking and three positive emotion categories (Content, Nostalgic and Tame/safe) (Figure 2.2). In the Lab context the two products were only differentiated on two emotion categories and there was no significant difference on liking. When participants tasted the beers in the Lab setting, the Ale was scored significantly higher than the Lager on Shocked and lower on Tame/Safe. In the Evoked context only one significant difference between the two beers was found, namely a significantly higher score for the Ale on Disgusted. Table 2.4 shows the number of participants that, given the choice, would take home either the Lager or the Ale in each of the three context conditions, together with the results from the associated Cochran's Q test. A significant majority of the participants chose to receive the Lager in the Bar ($p = 0.028$), while in the Lab ($p = 0.518$) and Evoked context ($p = 0.750$) there was no significant difference in the number of participants that chose to receive the Lager or the Ale (Table 2.4). These results showed that context had a significant effect on product choice and not surprisingly, a significant majority of the participants chose to receive their best-liked product and that was associated with a positive emotional response.

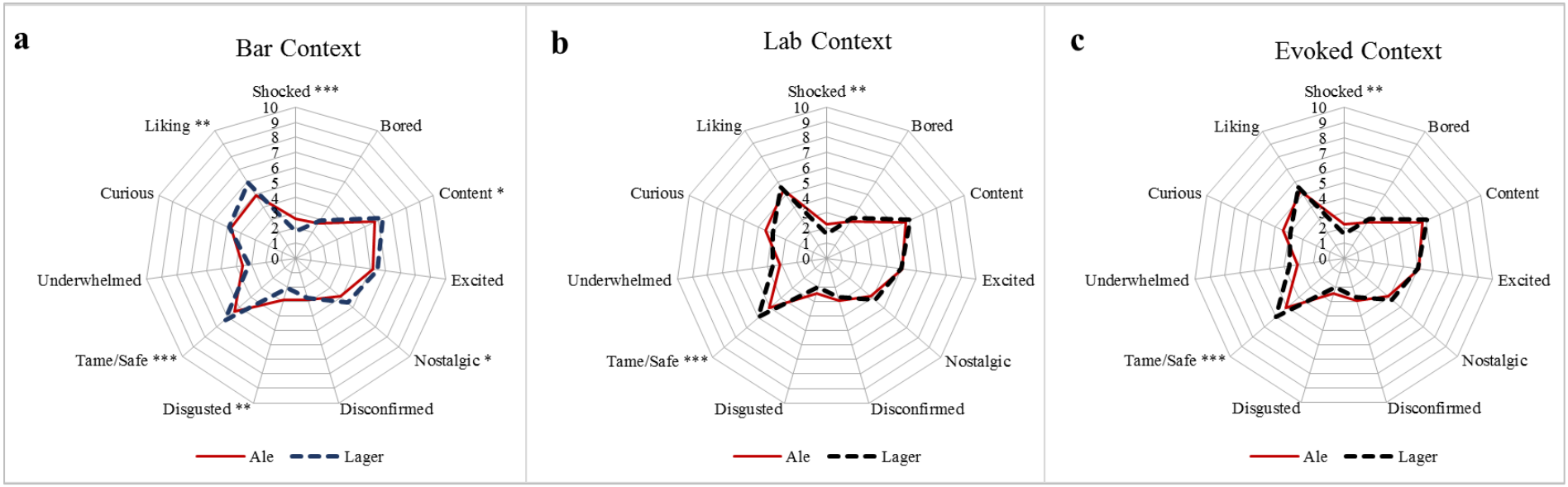


***Figure 2.2**: Mean scores for the ale and lager when evaluated by 100 participants in the Bar (a) the Lab (b) and the Evoked context (c). Terms with *** are significantly different between the two products at p < .001; ** p < .01; * at p < .05.*

Table 2.4: *Number of participants that chose to receive either the lager or the ale in each context (Bar, Lab, Evoked) and observed Q and p-value from Cochran's Q test per context condition (DF=1). Bold font highlights significant p-value at an α risk of 0.05. Participants that did not express a preference for either product were excluded from the analysis.*

	Bar	Lab	Evoked
Lager	56	46	46
Ale	35	40	43
Q (observed)	4.85	0.10	0.42
p-value	**0.028**	0.518	0.750

2.3.2 Response of different consumer clusters to beer in each context

2.3.2.1 Three clusters based on product liking per context

To study the relative effects of context versus hedonic preference on emotional response and product choice, participants were clustered based on their liking of each product in each context. The cluster analysis revealed three consumer clusters that differed in their overall liking for the Ale and the Lager in the three context conditions. The average scores for liking and emotional response per cluster are depicted in Figure 2.3a-f and listed in Table 2.5a-c. Table 2.6a-c displays the *F* and *p*-values from the four-factor ANOVAs per cluster for each emotion category and liking. One cluster containing 26 participants was characterised as rating the Ale significantly higher in liking than the Lager in each context (Table 2.5a and 2.6a). A second cluster, (34 participants) gave significantly higher liking scores for the Lager than for the Ale under each context condition (Table 2.5b and 2.6b). The third and largest cluster consisted of 40 participants that, unlike the other two clusters, did not show a consistent hedonic preference for one of the two products (Table 2.5c and 2.6c). The clusters were consequently named the ‘Ale Likers’, ‘Lager likers’ and the ‘Context Sensitives’ respectively.

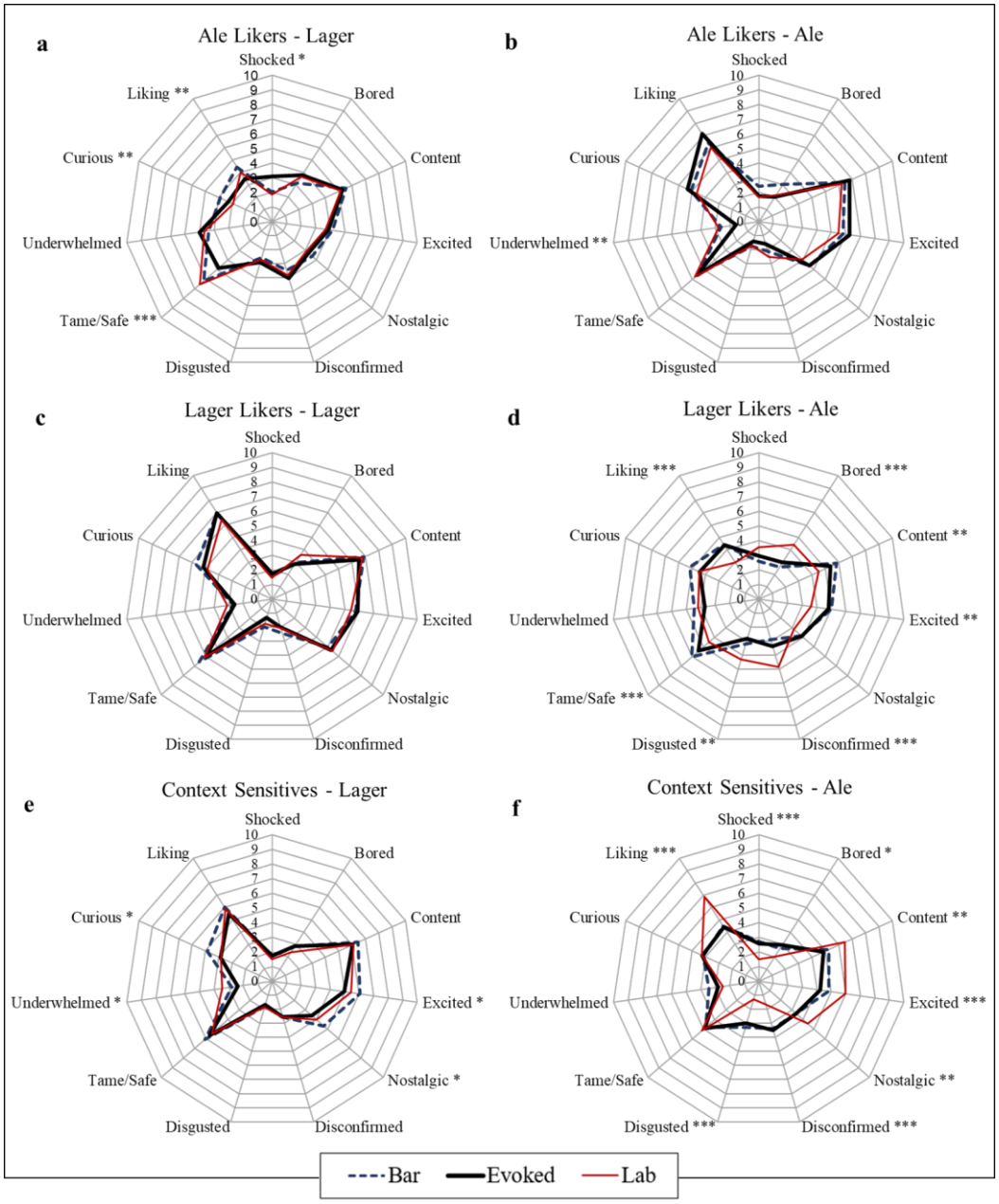


Figure 2.3*: Mean scores for the Lager (a, c, e) and Ale (b, d, f) when evaluated by each of three consumer clusters (Ale Likers, Lager Likers and Context Sensitives) in the Bar, the Lab and the Evoked context. Terms with *** are significantly different between the two products at p < .01; ** p < .05; * at p < .1.*

Table 2.5: *Mean scores and standard errors (SE) for the lager and ale when evaluated in the Bar, Lab and Evoked context by each of three consumer clusters (Ale Likers, Lager Likers and Context Sensitives). All data was collected on a continuous line scale from 0 to 10.*

a) Ale likers (N=26)

	Bar				Lab				Evoked			
	Lager		Ale		Lager		Ale		Lager		Ale	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Shocked	2.05	0.37	2.41	0.35	1.85	0.41	1.69	0.38	3.11	0.55	1.75	0.39
Bored	3.14	0.46	3.06	0.42	3.67	0.61	2.15	0.31	3.78	0.52	2.03	0.43
Content	5.52	0.42	6.47	0.36	5.13	0.27	6.22	0.37	5.27	0.47	6.82	0.39
Excited	4.15	0.41	5.78	0.41	3.67	0.32	5.48	0.41	3.89	0.42	6.26	0.34
Nostalgic	3.59	0.43	4.63	0.55	3.07	0.45	3.92	0.52	3.27	0.43	4.58	0.46
Disconfirmed	3.45	0.52	2.09	0.36	3.83	0.41	2.50	0.47	4.02	0.55	1.60	0.39
Disgusted	2.56	0.44	1.76	0.33	2.71	0.37	1.69	0.42	2.88	0.45	1.40	0.38
Tame/safe	6.18	0.40	5.48	0.44	6.52	0.39	5.75	0.49	4.84	0.50	5.58	0.58
Underwhelmed	4.39	0.56	2.62	0.42	4.68	0.51	2.78	0.42	5.02	0.63	1.58	0.26
Curious	3.91	0.48	5.09	0.44	2.92	0.43	4.65	0.53	3.30	0.50	5.34	0.48
Liking	4.42	0.39	6.40	0.43	4.01	0.30	6.03	0.49	3.46	0.32	7.15	0.30

Table 2.5 continued

b) Lager Likers (n=34)

	Bar				Lab				Evoked			
	Lager		Ale		Lager		Ale		Lager		Ale	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Shocked	1.51	0.23	2.56	0.46	1.44	0.25	3.55	0.52	1.74	0.30	2.92	0.48
Bored	2.94	0.35	2.61	0.33	3.58	0.39	4.38	0.41	2.82	0.39	2.96	0.38
Content	6.89	0.28	5.84	0.45	6.79	0.29	4.50	0.40	6.49	0.37	5.37	0.43
Excited	5.76	0.37	5.00	0.51	5.42	0.39	3.61	0.39	5.92	0.38	4.80	0.38
Nostalgic	5.04	0.45	3.87	0.48	5.44	0.43	3.16	0.38	5.33	0.42	3.88	0.38
Disconfirmed	2.44	0.38	2.90	0.47	2.08	0.34	4.84	0.46	2.01	0.31	3.36	0.44
Disgusted	1.98	0.32	3.23	0.48	1.75	0.33	4.32	0.48	1.33	0.23	2.82	0.46
Tame/safe	6.60	0.32	6.05	0.42	6.13	0.37	4.51	0.44	5.99	0.46	5.42	0.44
Underwhelmed	2.71	0.36	4.44	0.53	3.10	0.41	4.18	0.42	2.61	0.37	3.71	0.46
Curious	5.70	0.39	5.19	0.39	4.87	0.41	4.47	0.38	5.19	0.42	4.44	0.41
Liking	6.99	0.27	4.41	0.47	6.40	0.33	2.97	0.30	6.98	0.28	4.35	0.44

1641

1642

Table 2.5 continued

c) Context Sensitives (N=40)

	Bar				Lab				Evoked			
	Lager		Ale		Lager		Ale		Lager		Ale	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Shocked	1.73	0.23	2.71	0.41	1.50	0.26	1.50	0.27	1.73	0.32	2.59	0.43
Bored	2.85	0.33	2.68	0.35	2.38	0.29	1.98	0.28	2.84	0.37	2.94	0.35
Content	6.45	0.30	5.24	0.40	6.08	0.39	6.44	0.39	6.05	0.33	4.84	0.43
Excited	6.05	0.28	4.82	0.39	5.45	0.38	5.98	0.36	4.98	0.40	4.26	0.38
Nostalgic	4.66	0.37	3.37	0.35	4.03	0.41	4.44	0.46	3.60	0.43	3.37	0.43
Disconfirmed	2.57	0.36	3.37	0.44	2.54	0.39	1.66	0.30	2.52	0.34	3.48	0.44
Disgusted	1.80	0.26	3.29	0.44	1.84	0.36	1.31	0.23	1.68	0.29	2.97	0.44
Tame/safe	6.09	0.36	4.81	0.44	5.46	0.40	5.15	0.37	5.64	0.41	4.89	0.43
Underwhelmed	2.80	0.32	3.44	0.41	3.44	0.45	2.46	0.37	2.36	0.32	2.81	0.40
Curious	4.87	0.36	4.18	0.42	3.83	0.39	4.27	0.44	3.85	0.41	4.23	0.42
Liking	6.03	0.31	4.48	0.39	5.90	0.38	6.85	0.31	5.51	0.34	4.43	0.42

***Table 2.6**: F and p-values from a four-factor ANOVA (Participant, Context, Product, Session) per each of three consumer clusters (Ale Likers, Lager Likers and Context Sensitives). Bold font highlights significant p-values (p<0.05).*

a) Ale likers (N=26)

	Main effects 4-factor ANOVA								Interaction effects 4-factor ANOVA							
	Participant (DF=25)		Context (DF=2)		Product (DF=1)		Session (DF=2)		Context* Product		Context* Session		Product* Session		Context* Product* Session	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Shocked	2.13	**0.004**	2.04	0.135	1.55	0.216	2.741	0.069	2.63	0.077	0.96	0.433	0.043	0.958	0.32	0.862
Bored	4.33	< **0.001**	0.08	0.926	13.54	< **0.001**	1.967	0.145	2.80	0.065	2.80	**0.029**	1.059	0.350	0.77	0.546
Content	2.90	< **0.001**	0.58	0.560	19.90	< **0.001**	4.801	**0.010**	0.75	0.472	0.53	0.717	0.607	0.547	0.27	0.900
Excited	5.42	< **0.001**	1.14	0.322	66.53	< **0.001**	4.319	**0.016**	1.74	0.180	0.35	0.845	1.951	0.147	0.38	0.821
Nostalgic	7.07	< **0.001**	2.54	0.083	15.27	< **0.001**	2.328	0.102	0.37	0.692	0.56	0.691	1.100	0.336	0.99	0.415
Disconfirmed	2.33	**0.001**	0.48	0.620	26.41	< **0.001**	2.718	0.070	0.59	0.556	0.97	0.426	2.021	0.137	1.57	0.188
Disgusted	2.00	**0.007**	0.03	0.966	13.15	< **0.001**	1.525	0.222	0.37	0.695	1.70	0.154	0.029	0.972	0.86	0.489
Tame/safe	4.56	< **0.001**	3.39	**0.037**	0.64	0.425	1.039	0.357	2.59	0.079	0.44	0.778	0.401	0.670	0.30	0.875
Underwhelmed	3.61	< **0.001**	0.33	0.721	52.88	< **0.001**	4.698	**0.011**	2.57	0.081	0.52	0.722	0.034	0.966	0.38	0.821
Curious	4.94	< **0.001**	2.38	0.098	31.80	< **0.001**	1.598	0.207	1.12	0.331	3.40	**0.011**	0.624	0.537	1.12	0.352
Liking	3.72	< **0.001**	1.11	0.333	102.46	< **0.001**	4.061	**0.020**	4.23	**0.017**	0.30	0.877	0.655	0.521	1.60	0.179

Table 2.6 continued

b) Lager Likers (n=34)

	Results 4-factor ANOVA								Interaction effects 4-factor ANOVA							
	Participant (DF=33)		Context (DF=2)		Product (DF=1)		Session (DF=2)		Context* Product		Context* Session		Product* Session		Context* Product*Session	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Shocked	3.50	< **0.001**	0.95	0.388	30.40	< **0.001**	0.81	0.449	1.43	0.243	1.08	0.367	0.25	0.780	1.01	0.405
Bored	3.01	< **0.001**	7.74	**0.001**	0.57	0.453	0.42	0.661	1.36	0.259	0.59	0.669	0.25	0.777	0.14	0.965
Content	4.15	< **0.001**	2.60	0.078	35.71	< **0.001**	0.42	0.658	2.45	0.090	0.26	0.902	0.09	0.910	1.42	0.231
Excited	3.38	< **0.001**	3.80	**0.024**	19.37	< **0.001**	1.13	0.324	1.03	0.358	0.18	0.949	0.13	0.881	2.42	0.051
Nostalgic	5.04	< **0.001**	0.43	0.649	37.02	< **0.001**	0.04	0.960	1.48	0.230	1.10	0.358	0.09	0.914	1.67	0.159
Disconfirmed	2.57	< **0.001**	2.90	0.058	26.01	< **0.001**	0.56	0.575	4.76	**0.010**	0.56	0.690	0.04	0.957	0.24	0.917
Disgusted	3.76	< **0.001**	4.30	**0.015**	45.70	< **0.001**	1.87	0.157	2.02	0.136	0.96	0.433	0.53	0.591	1.73	0.145
Tame/safe	5.35	< **0.001**	4.94	**0.008**	13.63	< **0.001**	0.71	0.495	1.81	0.168	0.76	0.553	0.18	0.838	3.35	**0.012**
Underwhelmed	3.28	< **0.001**	1.08	0.343	19.50	< **0.001**	0.47	0.628	0.57	0.565	1.23	0.302	0.18	0.834	0.45	0.774
Curious	4.25	< **0.001**	2.41	0.093	4.50	**0.036**	3.03	0.051	0.15	0.861	2.03	0.093	0.06	0.937	1.18	0.320
Liking	3.62	< **0.001**	7.25	**0.001**	136.22	< **0.001**	0.09	0.913	1.28	0.281	0.20	0.940	0.02	0.977	0.20	0.940

Table 2.6 continued

c) Context Sensitives (N=40)

	Results 4-factor ANOVA								Interaction effects 4-factor ANOVA							
	Participant (DF=39)		Context (DF=2)		Product (DF=1)		Session (DF=2)		Context* Product		Context* Session		Product* Session		Context* Product* Session	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Shocked	4.29	< **0.001**	4.22	**0.016**	8.28	**0.004**	1.64	0.196	2.39	0.095	1.74	0.143	0.38	0.682	0.75	0.560
Bored	3.83	< **0.001**	2.65	0.074	0.53	0.468	7.38	**0.001**	0.56	0.573	0.65	0.628	0.24	0.789	2.14	0.078
Content	4.13	< **0.001**	3.10	**0.048**	8.12	**0.005**	2.40	**0.093**	4.61	**0.011**	1.91	0.110	0.08	0.921	0.71	0.584
Excited	4.22	< **0.001**	6.77	**0.001**	3.84	0.052	1.46	**0.235**	4.62	**0.011**	1.09	0.362	0.73	0.482	0.95	0.437
Nostalgic	4.02	< **0.001**	2.70	0.070	1.82	0.179	0.07	**0.934**	3.20	**0.043**	0.71	0.588	0.17	0.848	0.35	0.841
Disconfirmed	2.29	< **0.001**	4.22	**0.016**	1.09	0.298	4.53	**0.012**	5.12	**0.007**	1.22	0.302	0.99	0.372	0.61	0.657
Disgusted	2.42	< **0.001**	5.09	**0.007**	8.78	**0.003**	3.42	**0.035**	6.63	**0.002**	0.72	0.576	0.17	0.841	0.78	0.542
Tame/safe	7.84	< **0.001**	0.06	0.945	12.24	**0.001**	4.23	**0.016**	1.74	0.178	1.60	0.177	0.63	0.535	0.69	0.601
Underwhelmed	2.90	< **0.001**	1.72	0.182	0.02	0.894	0.82	**0.443**	3.33	**0.038**	0.81	0.519	0.00	0.998	0.39	0.818
Curious	3.00	< **0.001**	0.86	0.427	0.02	0.887	2.17	0.117	1.54	0.217	0.07	0.990	0.03	0.972	0.54	0.706
Liking	2.49	< **0.001**	9.10	<**0.001**	4.55	**0.034**	4.86	**0.009**	9.55	**0.000**	1.85	0.121	1.39	0.251	0.82	0.511

2.3.2.2 Ale likers

Within the Ale Likers cluster ANOVA revealed that this cluster scored the Ale significantly higher than the Lager on four out of five positive emotion categories (Content, Excited, Nostalgic and Curious) and significantly lower on four out of five negative emotion categories (Bored, Disconfirmed, Disgusted and Underwhelmed) (Table 2.5a and 2.6a). The emotional response of this cluster to the two products is visually represented by Figure 2.3a and 2.3b. The four-factor ANOVA revealed a significant context effect for the emotion category Tame/safe within the Ale Likers cluster (Table 2.6a). Comparing the responses between the three context conditions separately for each product revealed slightly more context variation for the Lager than the Ale within the Ale Liker cluster (Figure 2.3a and 2.3b). Session order effect was significant for Content, Excited, Underwhelmed and Liking. Tukey post hoc groupings showed a more positive response during the first session. ANOVA also revealed a significant interaction effect between context and product for liking (Table 2.6a), which was caused by a stronger product differentiation on liking in the Evoked context (Table 2.5a). Figure 2.4 illustrates the choice behaviour of each cluster in each context. As can be seen in Figure 2.4a most of the Ale Likers chose to receive the Ale in each of the three contexts. According to Cochran's Q analysis the Ale was chosen significantly more frequently than the Lager in the Bar (Q (1, n = 22) = 18.18, $p < 0.001$), Lab context (Q (1, n = 21) = 13.76, $p < 0.001$) and Evoked (Q (1, n = 25) = 21.16, $p < 0.001$) by this cluster.

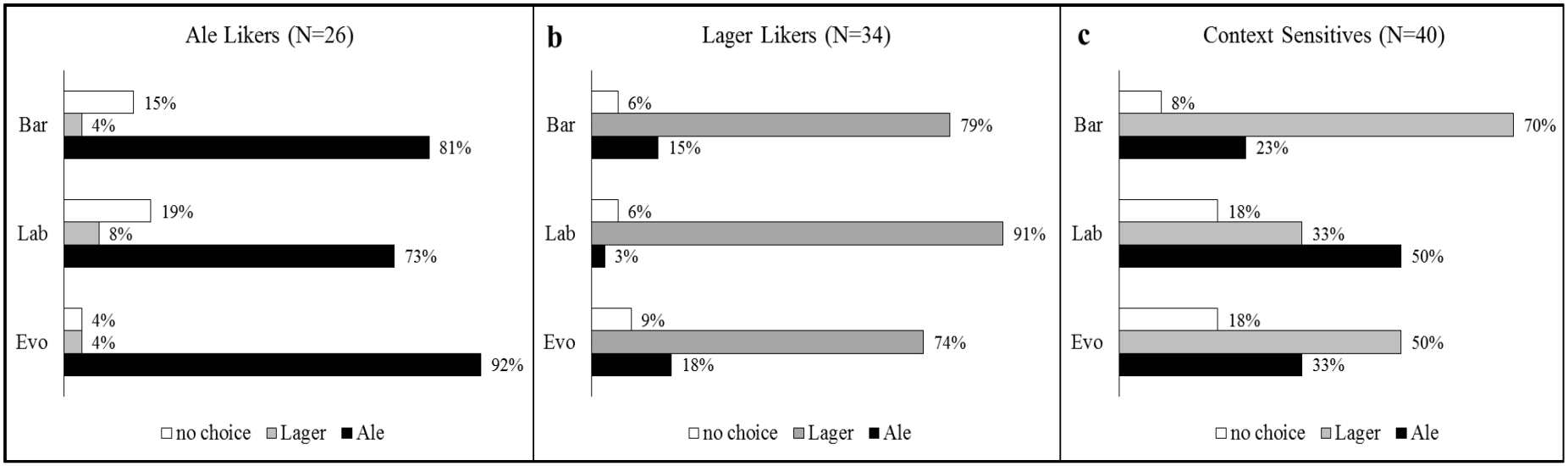


Figure 2.4: *Percentage of participants per cluster of Ale Likers (a) Lager Likers (b) and Context Sensitives (c) that chose to receive the ale, the lager or did not make a choice between the two products when tasting the products in the bar ('Bar'), the sensory laboratory ('Lab') and evoked context ('Evo')*

2.3.2.3 Lager Likers

The cluster of Lager Likers scored the Lager significantly higher than the Ale on all positive emotion categories (Content, Excited, Nostalgic, Tame/safe and Curious) and lower on four out of five negative emotion categories (Shocked, Disconfirmed, Disgusted and Underwhelmed) (Table 2.5b and 2.6b). The response of the Lager Likers to the Ale differed significantly between the context conditions on four out of ten emotion categories (Bored, Excited, Disgusted, Tame/safe) as well as on liking (Table 2.6b). For three more emotion categories context effect was approaching significance (Content, Disconfirmed and Curious) (Table 2.6b). No significant differences were found in how the Lager Likers felt about the Lager in the three different context conditions (Figure 2.3c). For the Ale however, six emotion categories (Bored, Content, Excited, Disconfirmed, Disgusted and Tame/Safe) and liking were rated differently depending on which context the Ale was tasted (Figure 2.3d). No significant session effects were found within this cluster. The four-factor ANOVA revealed a significant context*product interaction effect for Disconfirmed (Table 2.6b) which was a result of a significantly higher feeling of Disconfirmation for the Ale in the Lab than in the Bar and Evoked context (Figure 2.3d). A significant three-factor interaction between context, product and session was found for Tame/Safe (Table 2.6b). This significant effect was caused by relatively high scores for the Ale during the first session in the Bar context and relatively low scores for the Ale during the first session in the Lab. From Figure 2.4b it can be seen that in each context the majority of the Lager Likers chose to receive the Lager. Cochran's Q analysis revealed that the Lager was selected significantly more frequently than the Ale by the Lager Likers in

the Bar (Q (1, n = 32) = 15.13, $p < 0.001$), Evoked (Q (1, n = 31) = 11.65, p = 0.001) and Lab context (Q (1, n = 32) = 28.13, $p < 0.001$).

2.3.2.3 Context Sensitives

Unlike the Ale Likers and Lager Likers, the Context Sensitives did not have a stable hedonic preference for one of the two products. There were significant context, product and context*product interaction effects on liking for this cluster (Table 2.6c). The liking for the Lager did not vary significantly ($p < 0.05$) between the three contexts, although for Excited, Nostalgic, Underwhelmed and Curious context effects neared significance ($p < 0.10$) (Figure 2.3e). The Ale however, received a significantly higher liking score in the Lab than in the Bar and Evoked context (Figure 2.3f). For the emotional response ANOVA revealed significant contexts on five out of ten categories (Shocked, Content, Excited, Disconfirmed and Disgusted) (Table 2.6c), as well as significant interaction between product and context on six out of ten emotion categories (Content, Excited, Nostalgic, Disconfirmed, Disgusted and Underwhelmed) (Table 2.6c). What is striking from the results in Figure 2.3f is that in the Lab context, the Context Sensitives gave the Ale higher scores on positive emotion categories and lower scores for negative emotions compared to when they tasted the same product in the Bar or Evoked Context. Session order had an effect on eight out of ten emotion categories and liking. Examining the Tukey post hoc groupings made clear that this was due to a first order effect where participants reacted more positively to the beers during their first session. No significant interaction effects between context*session, product*session or context*product*session were found for this cluster.

The choice behaviour of the Context Sensitive cluster differed depending on the context condition (Figure 2.4c). In the Bar, Cochran's Q test showed that the Lager was chosen over the Ale by a significant number of participants in the Context Sensitive cluster (Q (1, n = 37) = 9.76, p = 0.002). In the Lab context however, there was no significant difference between the number of times the Lager and the Ale were chosen by the Context Sensitive cluster (Q (1, n = 33) = 1.48, p = 0.223). Visually, the choice behaviour in the Evoked context appears to approach the choice behaviour in the Bar, as Figure 2.4c shows a larger proportion of participants choosing the Lager in the evoked context. However, Cochran's Q analysis revealed that there was no significant difference between the number of times the Lager and the Ale were chosen by the Context Sensitive cluster in the Evoked context (Q (1, n = 33) = 1.48, p = 0.223).

2.3.2.4 Emotional Context-Product-Cluster Space

The emotional space for the two samples as evaluated by the three clusters in the three contexts was visualised using PCA. Figure 2.5 displays the bi-plot of the first two principal components that accounted for 80.28% of the variance in the data and represents the different positioning of the products for each cluster in the emotional space. Most of the variance (67.61%) was explained by the first principal component (PC1) which was positively correlated with emotion categories Disconfirmed (r (18) = 0.97), Disgusted (r (18) = 0.91), Shocked (r (18) = 0.82), Underwhelmed (r (18) = 0.80) and Bored (r (18) = 0.69). The emotion categories Content (r (18) = -0.96), Excited (r (18) = -0.95), Nostalgic (r (18) =-0.86), Curious (r (18) = -0.64) and Tame/safe (r (18) = -0.48) were negatively correlated with PC1. This component was described as being

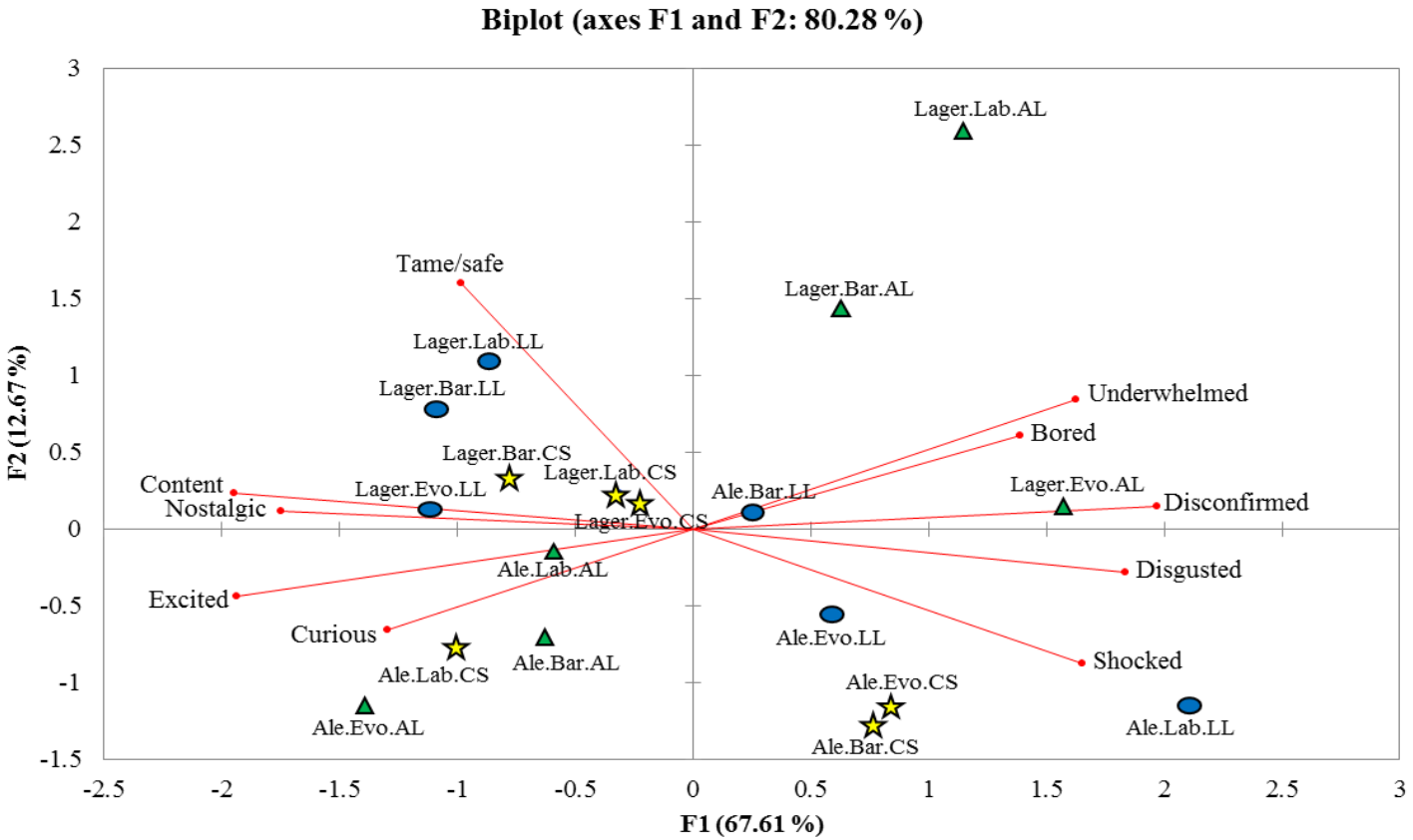


***Figure 2.5**: Bi-plot of the first two dimensions of the PCA showing the emotional space and positioning of the three clusters to the Ale (represented in Bold and Italic) and Lager (represented underlined) in each of the three context conditions, represented by Bar (Bar context), Lab (Sensory Laboratory) and Evo (Evoked context). The three clusters are represented by ● LL (Lager Likers), ▲ AL (Ale Likers) and ☆ CS (Context Sensitives).*

associated with pleasantness. The second principal component (PC2) explained an additional 12.67% of the variance in the data and was mainly driven by Tame/safe which correlated positively with PC2 (r (18) = 0.79). Underwhelmed (r (18) = 0.42) and Bored (r (18) = 0.30) also showed a positive correlation with F2. Similarly, Shocked (r (18) = -0.43), Curious (r (18) = -0.32) and Excited (r (18) = -0.22) negatively correlated with PC2. This component was described as being associated with level of activation. The emotional response to the Lager was projected towards the positive direction of PC2 for all three clusters, regardless of context. The emotional response to the Ale was mostly projected towards the negative direction of PC2, with the exception of the response of the Lager Likers to the Ale in the Bar which was located closer to the centre of the plot (Figure 2.5).

The emotional response of the Ale Likers to the Ale was projected towards the negative direction of both PC1 and PC2 in Figure 2.5, indicating that the emotional response of this cluster to the product they preferred was towards high activation and pleasantness. Their emotional response to the Lager was projected in the positive direction of PC1 and PC2 and hence associated with low activation and unpleasantness (Figure 2.5). For Ale Likers Figure 2.5 depicts an effect of context in the second dimension where the responses in the Lab were projected more towards a low level of activation than the responses from the Bar and the Evoked context for both products. This relates to the significant context effect that was found for Tame/safe for this cluster (Table 2.6a).

The emotional response of the Lager Likers to the Lager was projected in the negative direction of PC1 and positive direction of PC2, indicating that their

emotional response to the product they liked best was associated with low activation and pleasantness. The emotional response of the Lager Likers to Lager did not differ significantly between the context conditions (Figure 2.3c). To the Ale however, Lager Likers did respond differently depending on the context (Figure 2.3d). The emotional response of Lager Likers to the Ale in the Lab was associated with high activation and unpleasantness and consequently projected in the lower right corner of the bi-plot (Figure 2.5). The emotional response to the Ale in the Bar was located closer to the centre of the bi-plot and less far from the response of this cluster to the Lager (Figure 2.5). It appears that for the Lager Likers, tasting Ale in a Bar made their emotional response more similar to that of the Lager, which as a consequence, reduced product differentiation. In the Evoked context the Lager Likers' emotional response to the Ale was also associated with higher activation and unpleasantness, but to a lesser extent than in the Lab context (Figure 2.5). In fact, Evoking a context made their emotional response more similar to their response in the Bar compared to their response in the Lab.

The difference in emotional response of the Context Sensitives to the Ale in the Lab compared to the Bar and Evoked context can be clearly seen in Figure 2.5. While the emotional response of Context Sensitives to the Ale in the Bar and Evoked context was associated with high activation and unpleasant emotions, the response to the same product in the Lab was associated with high activation and pleasantness (Figure 2.5). The emotional response of the Context Sensitive cluster to the Lager was close to the centre of the bi-plot and associated with low activation and pleasantness (Figure 2.5). The location of the Bar responses in the emotional space show that the Context Sensitive

cluster behaves similarly to the Lager Likers when tasting the products in the Bar, but in the Lab their emotional response to the Ale becomes more positive and thus more similar to their response to the Lager (Figure 2.5), with a weaker product differentiation as a result. The responses in the Evoked context approach the results from the Bar for the Context Sensitive cluster (Figure 2.5), leading to a stronger product differentiation than in the Lab context.

2.4. Discussion

2.4.1 Context affects product differentiation on liking, emotions and choice

Although only one emotion category (Curious) appeared to be significantly impacted by context when all 100 participants were considered, analysing the data separately for each context condition revealed a clear difference between the Bar and Lab in terms of product differentiation on liking, emotional response and product choice. The stronger differentiation on liking in a natural context is in line with other studies that found that liking differentiated between products in a real life environment but not in the sensory lab (Boutrolle et al., 2007; Hersleth et al., 2005; Holthuysen et al., 2017). Previous research has suggested that context has an effect on product choice (De Graaf, Kramer, et al., 2005; Sester et al., 2013). In the current study, overall beer choice was in line with liking and a positive emotional response in the Bar context while in the Lab there was no significant difference in choice for either beer. It should be noted that even though effort was made to make the bar condition as 'natural' as possible by conducting the experiment in the evenings during normal opening hours, letting participants pick up their own beers at the bar, allowing free social interaction, and using participants' own smartphones to complete the questionnaire, there are limitations as to how this

experiment compares to a real life consumption situation. Participants were aware of the fact that they were part of an experiment and were asked to consciously evaluate and report their emotional response using a questionnaire. The use of direct questions has been criticised for inducing analytical attitudes that may disturb the natural relationship with the product (Köster, 2003). Secondly, the beer tasting protocol can be considered 'unnatural' since participants were only allowed to take one sip of each beer and asked to rinse their palate with water between the samples. In real life situations consumers are free to choose the product they wish to consume and having a choice has been shown to impact product acceptability (De Graaf, Cardello, et al., 2005). However, the participants in the current study did not get to choose the beers before tasting them. In addition, participants did not receive any information on which products they were tasting and as mentioned by Dijksterhuis (2016), consumers usually do not encounter a product without knowing anything about what it is they are consuming. These limitations to the realism of the Bar context also apply to the Evoked context. In addition it is unclear how much time participants actually spent watching the picture slideshow under the Evoked context condition and whether they made an effort to imagine being in the Bar. When analysing the results from all 100 participants together, evoking a bar context using pictures and sound recordings did not appear to modify results towards those obtained when participants were physically in the Bar environment. Other researchers, using different levels of context evocation, found that these methods led to stronger product differentiation on liking (Hein, Hamid, Jaeger, & Delahunty, 2012; Holthuysen et al., 2017). The current study found no significant effect of evoking context on liking or

product choice when considering all participants. In addition, only one emotion category differentiated between the two products when the Evoked context was applied. More immersive methods of context evocation could possibly have a greater effect on consumer responses than the method used in this study.

Although product differentiation on emotional response was stronger in the Bar, it should be noted that the direction of the product differences was the same in all three contexts, namely a preference for the Lager, which was expressed by higher scores for positive emotion categories and lower scores for negative emotion categories compared to the Ale. This finding shows that overall preference was not changed by the different contexts, only the degree of product differentiation.

2.4.2 Emotional response provides information beyond liking and relates to product choice

The results of this study agree with previous studies that products can be differentiated by measuring consumer emotional response with questionnaires (Beyts et al., 2017; King & Meiselman, 2010; Ng et al., 2013a). Products' sensory attributes are a source of food emotions (Desmet & Schifferstein, 2008) and measuring emotional responses to products provides incremental information to overall acceptability data (King et al., 2010). When comparing the relative effectiveness of liking, emotional response and product choice measures to discriminate between products in this study, it is clear that in the Lab and Evoked context emotional response allowed for product differentiation while there was no significant difference in liking or product choice. This demonstrates the benefit of measuring emotional response for further discriminating between products that have similar hedonic ratings. It

has been argued that a single hedonic measurement is insufficient for the prediction of the acceptance of products in the long term (Köster, 2003). Similarly, it would be of interest to study the stability of emotional responses over time and after repeated exposure to a given product.

2.4.3 Clustering participants provides valuable insights

Looking at the cluster results it is clear that a deeper insight regarding consumer responses and context was gained. Two consumer clusters that were discovered in the current study showed a clear hedonic preference for one of the two products that was stable over context. Regardless of whether the Lager Likers and the Ale Likers were tasting the products in the Bar, Lab or Evoked context condition, they expressed the same preference through their liking scores, emotional response and choice behaviour. This finding suggests that the hedonic preference, emotional response and choice behaviour of the Ale Likers and Lager Likers were driven by the sensory characteristics of the products rather than by the context in which they were exposed to the products. A third and largest cluster however, was found to be very sensitive to context. The effect that context had on the responses of the Context Sensitives demonstrates that for some consumers context can be just as important for preference as sensory characteristics, a fact that was earlier demonstrated by Hersleth, Mevik, Naes, and Guinard (2003). This leads to questions on whether performing consumer tests in natural consumption environments provides more reliable data than testing in sensory laboratories. More work needs to be done on the stability and reproducibility of data in real life consumption environments, where there is a lack of control over external sensory cues that might affect emotional response. Interestingly, for the Context Sensitives the

emotional response and choice behaviour in the Evoked context appeared to approach the results from the Bar, showing that for this cluster evoking a context led to results more similar to those of a real life consumption environment. This fact was lost when data was analysed with all participants included in the analysis together. Heterogeneity in consumer data in terms of emotional response to products has been shown before (Piqueras-Fiszman & Jaeger, 2016). Individual variation between consumers is apparent and treating consumers as a homogenous group is a common fallacy in sensory and consumer research (Köster, 2003). The current study shows the importance of clustering when looking at emotional response as well. It should be noted however, that the number of participants in this study (N = 100) was limited and therefore it would be interesting to see if similar patterns appear in similar studies with larger participant numbers.

2.4.4 Consumers differentiate beer on level of arousal

Similar to the results found by earlier studies on beverages (Chaya et al., 2015; Gutjar, De Graaf, et al., 2015; Ng et al., 2013a) the emotional space as visualised by the PCA was consistent with circumplex models of emotion that describe human emotion as a two-dimensional structure of pleasantness and activation (Larsen & Diener, 1992), with PC1 associated with pleasantness and PC2 with activation. Although the Ale Likers and Lager Likers had an opposite hedonic preference, the response to the product they scored highest on liking was similarly characterised by relatively high scores for positive emotion categories and low scores for negative emotion categories. It is striking that all three clusters differentiated the two products in the same way with regards to the activation dimension of the emotional space, with Lager

towards low activation emotional response and Ale towards the high activation emotional response. Silva et al. (2016) found that beer in general was associated with positive high arousal emotional responses. The current study shows that within the beer product category, some products can also be associated with relatively low activation emotional response, which is similar to findings by Chaya et al. (2015). These findings further demonstrate the benefit of measuring consumer emotional responses to beer products by providing additional information on consumer reactions to products beyond hedonic preference. The relationship between liking, emotional response and choice that was observed when considering all 100 participants together was seen at a cluster level as well. A significant majority of the Ale Likers and Lager Likers chose the product that they scored significantly higher on liking and positive emotion categories and lower on negative emotion categories, confirming previous research that has shown that liking and emotional response predict choice behaviour (Gutjar, De Graaf, et al., 2015).

2.4.5 Context effects are related to product liking

Similar to previous findings from other studies, the results from this experiment showed that emotional response to beer products is related to liking (Cardello et al., 2012; Chaya et al., 2015; Dorado, Chaya, et al., 2016; Silva et al., 2017) and it is not surprising that clustering participants on liking resulted in the discovery that the different consumer clusters differed in their emotional response to the products. The differences between clusters with different hedonic preference in terms of product-specific emotional response has been shown before with coffee (Bhumiratana et al., 2014) and wine (Jiang et al., 2016). Although the preference of the Ale Likers and Lager Likers was stable

over the three contexts, there were some differences between their responses from the Bar, Lab and Evoked context that could be observed from the results. Lager likers had a stronger negative emotional response to the Ale in the Lab, which was expressed by higher scores on negative emotion categories and lower scores on positive emotion categories compared to the Bar and Evoked context. Evoking a context seemed to assimilate the emotional response towards the Bar for the Lager Likers, mainly for the Ale, which was the product that they liked least. Ale likers responded to the products with a higher level of activation in their emotion response in the Evoked context, which was clearly visualised in the emotional space. Amongst the Lager Likers and Ale Likers more context effects were observed for products that were less liked than for products that were preferred by the individual clusters respectively. The results suggest that emotional response was more stable across contexts when a product was well liked. This offers a possible explanation as to why hedonic scores of some products and product categories are more affected by context than others (Boutrolle et al., 2007; De Graaf, Cardello, et al., 2005; Meiselman et al., 2000).

2.5. Conclusion

This study aimed to compare liking, emotional response and product choice of beer products in a natural consumption context, i.e. a bar, with a traditional sensory test setting, and to determine the relative effect of evoking a context. When all consumers were considered only limited effects of context on consumer response to beer products were found with no overall impact of context on liking. Overall, more product differentiation was evident in the bar than the other contexts, and emotions were shown to be more discriminating

than liking. However, analysing the data separately for three distinct consumer clusters of liking behaviour gave far deeper insights to the effect of context on consumer preference and the relative effects of evoking a context, highlighting the importance of consumer segmentation. Two clusters had opposing but stable product preference no matter what the context, but interestingly, the product preference of the third and largest cluster changed depending on the context, preferring the Lager in the bar but showing no preference in the lab. This highlights the importance of carrying out consumer testing of beer, and potentially other products, in realistic contexts. Evoking a bar context may provide a more efficient alternative but as it produced different effects for the different clusters more research is recommended to understand the importance of elements within an evoked context to simulate a realistic environment.

Acknowledgements

The Authors wish to thank and acknowledge Hal MacFie, Hal MacFie Training Services, UK, for his input on the data analysis and subsequent interesting discussions. A special thanks to Imogen Ramsey for her help conducting the study and the staff at Mooch Student Union Bar at the University of Nottingham for their collaboration and support.

M. Nijman gratefully acknowledges The University of Nottingham and Anheuser-Busch InBev for funding this PhD research.

Chapter 3

3. The stability of self-reported emotional response and liking of beer: comparing a bar to a central location test and a single sip versus a glass of beer

Preliminary thoughts Chapter 3

The effect of test environment on emotional response data was clearly demonstrated in the first experiment discussed in Chapter 2. In addition, it was evident that there was variation among participants, both in terms of their emotional responses to beer depending on preference and the extent to which their emotional responses varied between consumption environments. These findings were based on single sip testing however, and thus required further investigation into the effects of consumed amount of beer on emotional response. Moreover, the lack of experimental control in the bar environment led to questions regarding the reliability of data collected in this environment. If the data collected in realistic consumption situations was highly unstable, it would not be advisable to test in this setting. Therefore the next experiment focussed on the stability of emotional response over the course of consuming a glass of beer and after repeated exposure in a controlled setting and natural consumption environment.

The stability of self-reported emotional response and liking of beer: comparing a bar to a central location test and a single sip versus a glass of beer

Marit Nijman[1], Rebecca Ford[1]

[1]*Sensory Science Centre, Division of Food, Nutrition and Dietetics, University of Nottingham, Sutton Bonington Campus, United Kingdom*

Key words: Context, Emotional response, Consumption, Reliability, Reproducibility, Liking, Beer

Highlights:

- Emotional responses varied between a bar and central location test
- Some emotions varied across the consumption of a single sip and a glass of beer
- Test-retest reliability of emotional response to beer was relatively consistent regardless of test environment.
- Consuming a glass of beer led to higher test-retest reliability of emotion measurement than taking a single sip.
- Including a training session prior to evaluation resulted in a more stable emotional response over subsequent sessions.

3.0 Abstract

Controlled settings for sensory evaluation are believed to offer reliability of results, yet lack ecological validity, but it is unknown whether emotional response to products is consistent. The current study investigated I) impact of context on emotional response from a single sip to an entire glass of beer, II) test-retest reliability of responses in a Bar and in a controlled setting, and III) effects of a training session prior to evaluation on the reliability of responses in subsequent sessions. Beer consumers (N=97) evaluated emotional response and liking to beer (3.5% ABV) after one sip, half a glass (≈142 ml) and glass (≈284 ml), using a self-report questionnaire in a real life context (i.e. Bar) and central location test (CLT). A subset of 62 participants repeated the session in each context in order to determine test-retest reliability with intra-class correlation (ICC). Prior to the test sessions emotional response and liking were measured for the same beer during a training session. ANOVA revealed that context impacted emotional response overall ($p < 0.1$), but not the change in response from first sip to a glass of beer ($p > 0.1$). Overall, participants felt more Content after a glass of beer than after a single sip (F (2) = 5.66, $p <$ 0.01). Test-retest reliability was comparable between Bar (ICC's 0.467 – 0.837) and CLT (ICC's 0.180 – 0.796), but generally higher for responses after a glass of beer than one sip (F (2) = 3.1, p = 0.053). Content and Excited were higher during the training session, but responses remained stable over subsequent sessions. These findings support the recommendation to include realistic contexts when measuring emotional response in sensory consumer studies, both in terms of test setting and consumed amount of beer.

3.1. Introduction

The limited ability of single hedonic ratings to predict long term food choices has led the field of sensory science to explore alternative strategies for sensory consumer research (Jaeger & Porcherot, 2017; Kaneko et al., 2018). Consumers' emotional responses to products have been shown to differentiate between products (King et al., 2013) and predict food choice more accurately than liking alone (Dalenberg et al., 2014). Even though more realistic contexts are recommended for consumer research to optimise the ability of test results to predict consumer behaviour in real life situations (De Graaf, Cardello, et al., 2005; Köster, 2003), the majority of sensory consumer research still takes place in controlled settings such as sensory laboratories or central location tests (CLT) (Galiñanes Plaza et al., 2019). Controlled environments are believed to offer reliability and robustness of results (Galiñanes Plaza et al., 2019). The decision to test products in either a controlled setting or a natural consumption environment can therefore be seen as a trade-off between reliability and ecological validity. Reliability refers to the consistency of the test (Tavakol & Dennick, 2011) and is commonly assessed with test-retest procedures, which use Pearson correlation (Cardello et al., 2012) or Intra-class correlation (Yen & Lo, 2002) to determine if repeating a test yields the same result. Ecological validity is commonly used to indicate whether the environment in which a test takes place is representative of the context of interest (Delarue et al., 2019; Galiñanes Plaza et al., 2019).

A recent study by Nijman et al. (2019) found evidence that self-report emotional response to beer can be significantly impacted by context. The emotional response that a cluster of context-sensitive consumers reported in a

bar differentiated clearly between two beers, whilst their responses did not differ between the same two beers when measured in a sensory laboratory. The increased product differentiation found in a real life environment by Nijman et al. (2019) was based on measurements taken after consumption of a single sip of beer. Consumers would normally consume a whole glass of beer in real life consumption situations and thus a single sip may not accurately reflect true consumer experience and therefore ecological validity, especially as beer has a dynamic sensory and emotional profile over time and number of sips (Silva et al., 2019).

Measuring consumer responses in real life consumption situations can increase ecological validity, but at the same time an uncontrolled environment introduces noise variables (Galiñanes Plaza et al., 2019). In addition, instances of emotion can vary across individuals, as well as within a person across contexts (Feldman Barrett, 2016; Nijman et al., 2019). Therefore, it is unknown if a real life consumption situation (realistic context) could impact consumer experience from the first sip to finishing an entire glass of beer. Consequently, there is a need to determine whether the test environment impacts changes in emotional response to beer over time and number of sips.

Furthermore, emotional response may vary within an individual to the same product on different consumption occasions (Coppin & Sander, 2016; Feldman Barrett, 2016). The uncontrolled variables in a real life test setting are likely to vary between consumption occasions and could thus reduce test-retest reliability. Controlled sensory test environments are designed to keep external sensory cues to a minimum with the assumption that avoiding external noise facilitates reliable test results. Cardello et al. (2012) studied the test-retest

reliability of measuring consumers' emotional responses to food names and tasted foods in a controlled setting. They found high correlations for food names when measured at two instances, one week apart. For tasted foods the reliability was less stable. This was thought to be due to food names reflecting stable and broader cognitive images and associations, whereas tasted foods vary over time due to perceptual variability, changing expectations, and preceding appetitive contexts (Cardello et al., 2012). Moreover, Piqueras-Fiszman and Jaeger (2014a) found that emotional associations to food images under evoked contexts yielded stable check-all-that-apply responses when retested after one hour. These two studies suggest that emotional responses to food concepts and images are stable in controlled settings. However, it is unclear if emotional responses to beer consumption can be replicated in realistic contexts on separate consumption occasions.

Repeated exposure to the same product can change hedonic responses over time (Köster, Couronne, Léon, Lévy, & Marcelino, 2003) which may be due to increased familiarity, which can increase product liking, or loss of curiosity and rising product-boredom, which can reduce liking with time (Köster et al., 2003). It is currently unknown whether repeated exposure has an effect on emotional response to beer, although it appears that emotional responses to beer are more positive during the first session of a within-subject design than during subsequent sessions (Nijman et al., 2019). Using a 'dummy' sample has been shown to reduce first-order sample effects on emotional response to beer (Dorado, Pérez-Hugalde, et al., 2016). It would therefore be interesting to explore the effects of including a training session as a 'dummy' session, on

subsequent repeated exposure to the same beer on consumers' emotional response.

The current study had three objectives, namely to: I) determine the impact of context on emotional response to beer from the first sip to finishing an entire glass of beer, II) determine test-retest reliability of emotional response in a controlled setting and real life environment, III) explore the effects of including a training/dummy session on the emotional response to subsequent repeated exposure to beer.

3.2. Materials and methods

3.2.1 Experimental design

Beer consumers (N=97) evaluated their emotional response three times over the course of drinking a glass of beer (after first sip, after half a glass (142 ml) and after a full glass of beer (284 ml), using a self-report questionnaire under two context conditions; in a real life context (i.e. Bar) and in a central location test (CLT). Prior to the test sessions all participants completed a training session. A subset of 62 participants volunteered to perform the same session twice in each context in order to allow an evaluation of test-retest reliability. The study was approved by the Faculty of Medicine & Health Sciences Research Ethics Committee of the University of Nottingham (Ethics reference number 123-1709).



Figure 3.1*: On-campus student union bar used for Bar condition (left) and computer room used for Central Location Test (right) at the University of Nottingham.*

3.2.2 Participants

Participants were recruited among University of Nottingham undergraduate students (60%), postgraduate students (28%), staff (10%) and local volunteers external to the university (2%). The 97 participants (37% female) ranged in age between 18 and 62 with an average age of 23 (± 6.3) years. Native English speakers made up 84.5% (n = 82) of the participants. Non-native English speakers were given extra time to familiarise themselves with the English questionnaire and to ask questions during the training session. All participants drank beer at least once every two months and 73% stated they drank beer weekly. Individuals that had any reason to refrain from drinking beer (including self-declared health, pregnancy, religion or addiction) were excluded from participation. Participants received an inconvenience allowance for their participation.

3.2.3 Beer

A commercial lager beer (3.5% alcohol by volume) was used as the sensory stimulus in this study. All beer was from the same production batch. Sample preparation commenced once participants were seated and logged into the questionnaire. For each participant approximately 284 ml (half pint) of refrigerated (9°C ± 2) beer was poured from 500 ml cans into a standard unlabelled half pint glass (284 ml content). Serving temperature of the beer was 9.3°C (± 2.2).

3.2.4 Context conditions

The on-campus Student Union Bar at the University of Nottingham (Figure 3.1) was used as the Bar context. During experimental sessions the bar was open for business as usual and thus there was no control over external

variables such as music, room temperature and the presence of other customers that were not taking part in the experiment. Additionally, food was prepared and served in the bar which added food smells as an additional uncontrolled variable. At the start of the session participants received instructions, after which they were given a token that they could use to pick up their beer sample at the bar. Participants were free to sit anywhere in the bar and were asked to complete the questionnaire on their smartphone. Tablet devices were available for participants that did not have a smartphone. Interaction with other people was allowed, but participants were instructed not to talk about the beer or the questionnaire. The average room temperature in the bar was 21°C ±1.2.

The central location test (CLT) took place in a computer class room at the University of Nottingham (Figure 3.1). Participants were isolated from each other using portable booth dividers and instructed to complete the session in silence. Questionnaires were completed on individual desktop computers within each booth. Participants were not allowed to use their phone during the CLT sessions. Beer samples were distributed by the researcher and noise was minimised. Room temperature was monitored and averaged 20°C ±0.8.

3.2.5 Sessions

In advance of the experiment a 45-minute 'training session' was performed during which participants were familiarised with the emotion questionnaire and test protocol by consuming the beer and completing the questionnaire for the first time in the CLT location. The aim of the training session was to determine if a first exposure to the beer and the experimental situation would counteract first-order session effects, similar to how a 'dummy' sample can be used to neutralise first-order sample effects (Dorado, Pérez-Hugalde, et al.,

2016). In order to be able to investigate the effects of subsequent repeated exposure, the same beer was used for the training session as during the actual test sessions.

Participants' availability determined the order in which they were exposed to the two conditions after the training session. Forty-two participants (43%) completed the session in the Bar context first and 55 participants (57%) started in the CTL context. To determine test-retest reliability a subset of 62 participants volunteered to complete the session again in each context after completing the training and their first two sessions. Of this group, 35 participants (56%) started in the Bar context while the other 27 (44%) completed the CLT session first. All sessions were scheduled as 30-minute timeslots between 5pm and 9pm on Mondays, Tuesdays, Wednesdays and Thursdays. Participants were encouraged to come in at the same time and on the same day of the week for all sessions, as far as their schedules allowed. Depending on participants' availability there were one or two weeks between sessions, as well as between test and re-test.

3.2.6 Emotion questionnaire

Data was collected using an online questionnaire via Compusense Cloud (Version 8.0.6288.23054, Compusense Inc., Guelph, Ontario, Canada). During each session, participants reported their emotional response and liking for the beer (I) after the first sip, (II) after consuming half of the beer in the glass (approximately 142 ml) and (III) after consuming all of the beer in the glass (approximately 284 ml). A reduced beer specific emotion lexicon (Eaton et al., 2019) was used to determine emotional response at each consumption stage (Table 3.1). The emotion terms for each category were presented together and

participants were asked to rate each category on ten separate continuous line-scales of which the ends were labelled from ‘very low’ to ‘very high’ at 5% and 95% of the scale. Overall liking was rated after emotional response on a continuous line scale anchored from ‘dislike extremely’ at 5% to ‘like extremely’ at 95%.

Table 3.1: *Emotion categories and associated terms*

Shocked	Shocked, alarmed, cheated, confused, overwhelmed, strange, weird
Bored	Bored
Content	Content, calm, comfortable, comforted, enjoyment, good, happy, nice, pleasant, pleased, relaxed, satisfied
Excited	Excited, enthusiastic, fulfilled, fun, impressed, interested, optimistic, pleasantly surprised, want, warm
Nostalgic	Nostalgic, delirious, relieved
Disconfirmed	Disappointed, dissatisfied, unpleasantly surprised
Disgusted	Disgusted, horrible, repulsed, repelled, unpleasant
Tame/safe	Tame, safe
Underwhelmed	Underwhelmed
Curious	Curious

3.2.7 Data analysis

Statistical analyses were performed using Microsoft Excel 2013/XLSTAT (XLSTAT Version18.07.39020, Addinsoft, NewYork, USA). To determine the impact of context on emotional response and liking from the first sip through to finishing an entire glass of beer, mean scores and standard errors (SE) were calculated per consumed amount in each context, over the 97 participants that completed the first two test sessions. A two-factor ANOVA with context and consumed amount of beer as fixed factors and Fisher’s Least Significant Difference (LSD) post hoc test was performed to determine effects on emotional response and liking.

Test-retest reliability was determined using the data collected from the 62 participants that completed both the test session and the retest session in the Bar and CLT context. Intra-class Correlation Coefficients (ICC) were calculated for responses per emotion category and liking at each consumed amount in each context. ICC estimates were calculated using SPSS statistical package version 24 (SPSS Inc, Chicago, IL) based on a mean-rating (k= 2), absolute-agreement, 2-way mixed-effects model. In order to compare the effect of consumed amount on the test-retest reliability, the ICC's were averaged over variables (emotions and liking) per consumed amount in the Bar and CLT. A two-factor ANOVA was performed on the ICC's, with consumed amount and context as fixed factors.

The efficiency of using the training session as a 'Dummy' session was studied using the data collected from the 62 participants, by calculating the impact of this first session (training session) on the reproducibility of emotional response collected over the four subsequent test sessions (test sessions and retest sessions). For each participant the sessions were coded for chronologic session order, based on the date on which the participant took part in the session. The training session was coded as the first session for all participants, followed by the test sessions in the Bar and CLT context as either second or third, depending on individual session order. The re-test sessions were coded fourth and fifth. Mean scores and standard errors (SE) for emotional response and liking were calculated per session. The effect of session order on the responses were examined with a one-way ANOVA and Fisher's LSD post hoc test using session as factor.

3.3. Results

3.3.1 Impact of context on responses from first sip to entire glass of beer

Table 3.2 displays the *F* and *p*-values from the two-factor ANOVA comparing the two contexts (Bar and CLT) and consumed amount of beer (one sip, half a glass and an entire glass), as well as the interaction between context and amount. Context had an effect ($p < 0.05$) on the negative emotion category Shocked and positive emotion categories Content and Excited (Table 3.2). As can be seen from the means in Table 3.3, participants reported higher ratings for Shocked, Content and Excited in the Bar than in the CLT. In addition, Nostalgic ($p = 0.07$) and Disgusted ($p = 0.09$) appeared to be scores somewhat higher in the Bar than in the CLT (Table 3.3).

The amount that was consumed had an effect ($p < 0.05$) on Content (Table 3.2). As shown in Figure 3.2, post hoc HSD revealed the participants felt more Content after drinking a glass of beer than after a single sip or half the glass. In addition, participants appeared to report slightly higher scores on Curious ($p = 0.08$, Table 3.2) after one sip (Mean = 4.5, SE = 0.16) than after consuming half a glass (Mean = 3.9, SE = 0.16) or an entire glass of beer (Mean = 4.0, SE = 0.17).

No interaction effects were found between context and consumed amount ($p > 0.1$ or all variables, Table 3.2), indicating that the test environment did not impact changes in emotional response to beer over time and number of sips. Liking did not vary between the Bar and CLT context, nor over the course of consuming a single sip up to a glass of beer.

***Table 3.2**: F* and *p*-values from a two-factor ANOVA on emotional response and liking scores from 97 participants with context (Bar or CLT) and amount (sip, half glass & entire glass) as fixed factors. Bold font highlights *p*-values < 0. 05

		Context (DF=1)		Amount (DF=2)		Context*Amount (DF=2)	
		F	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	5.95	**0.01**	0.31	0.74	0.46	0.63
	Bored	0.51	0.48	1.38	0.25	1.28	0.28
	Content	7.79	**0.01**	5.66	**< 0.01**	0.27	0.76
	Excited	11.35	**< 0.01**	1.10	0.33	0.79	0.45
	Nostalgic	3.23	0.07	0.07	0.93	0.30	0.74
	Disconfirmed	0.85	0.36	0.55	0.58	1.05	0.35
	Disgusted	2.87	0.09	0.19	0.83	0.96	0.38
	Tame/Safe	0.84	0.36	0.06	0.94	0.48	0.62
	Underwhelmed	0.10	0.75	2.02	0.13	0.19	0.82
	Curious	2.12	0.15	2.59	0.08	0.18	0.84
	Liking	0.17	0.68	0.12	0.88	0.41	0.67

***Table 3.3**: Mean scores and standard errors (SE) for emotional response and liking for beer when evaluated by 97 participants in the Bar and Central location Test (CLT) after one sip (Sip), after consuming half of the beer in the glass (half) and after consumption of an entire glass of beer (Glass). All data was collected on a continuous line scale from 0 to 10.*

		Bar						CLT					
		Sip		Half		Glass		Sip		Half		Glass	
		Mean	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
Emotional response	Shocked	2.07	0.17	1.85	0.17	1.80	0.17	1.56	0.17	1.54	0.16	1.61	0.18
	Bored	3.37	0.21	2.87	0.21	2.70	0.22	2.88	0.21	2.80	0.21	2.90	0.24
	Content	6.14	0.17	6.48	0.15	6.81	0.18	5.86	0.19	5.94	0.19	6.39	0.21
	Excited	5.52	0.21	5.86	0.18	6.05	0.19	5.20	0.20	5.34	0.21	5.22	0.23
	Nostalgic	3.93	0.22	4.13	0.23	4.02	0.23	3.79	0.26	3.65	0.24	3.58	0.24
	Disconfirmed	2.39	0.19	2.17	0.19	1.96	0.16	2.04	0.19	1.92	0.20	2.13	0.22
	Disgusted	1.93	0.17	1.95	0.18	1.79	0.16	1.48	0.16	1.64	0.18	1.81	0.21
	Tame/Safe	6.06	0.21	5.94	0.21	6.23	0.21	5.91	0.22	5.98	0.22	5.85	0.23
	Underwhelmed	3.03	0.23	2.62	0.22	2.48	0.22	2.82	0.23	2.69	0.23	2.46	0.23
	Curious	4.52	0.21	4.18	0.22	4.18	0.23	4.40	0.23	3.78	0.24	3.88	0.25
	Liking	5.68	0.18	5.83	0.18	5.76	0.20	5.82	0.17	5.68	0.20	5.58	0.25

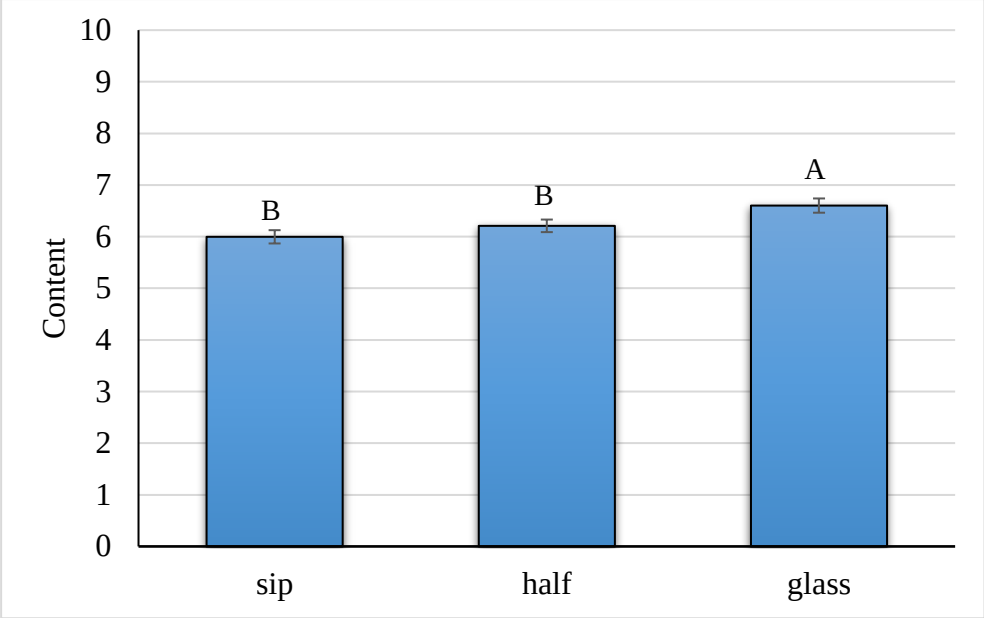


Figure 3.2*: Mean scores with standard errors (SE) for emotion category Content when the beer was evaluated by 97 participants after one sip (Sip), after consuming half of the beer in the glass (half) and after consumption of an entire glass of beer (Glass). A and B indicate Fisher's LSD post hoc groupings at p <0.05.*

3.3.2 Test-retest reliability of response per amount in the Bar and CLT

Table 3.4 shows the Intraclass Correlation Coefficients (ICC) for test-retest reliability of each emotion category and liking, evaluated per consumed amount in the Bar and CLT context by 62 participants. All of the reliability coefficients were positive (ICC +0.180 - +0.837) and most indicated a moderate (ICC +0.5 - +0.75) to good (ICC +0.75 - +0.9) test-retest reliability (Koo & Li, 2016).

The test-retest reliability was comparable between the two contexts, with a slight trend towards slightly higher ICC's in the Bar than the CLT ($F(1) = 3.9$, $p = 0.054$), which can also be seen from the averaged ICC's per amount and context in Figure 3.3.

***Table 3.4**: Intra Class Correlation coefficients between test and retest data collected from 62 participants at three stages of consumption (after a sip, half a glass and an entire glass) in the Bar and CLT. Cells are colour-scaled so that darker cells highlight higher coefficients.*

		Bar			CLT		
		1 sip	Half	Glass	1 sip	Half	Glass
Emotional response	Shocked	0.637	0.690	0.837	0.634	0.646	0.682
	Bored	0.452	0.629	0.640	0.605	0.473	0.765
	Content	0.596	0.554	0.676	0.364	0.180	0.532
	Excited	0.659	0.613	0.529	0.446	0.231	0.765
	Nostalgic	0.679	0.743	0.768	0.610	0.631	0.651
	Disconfirmed	0.467	0.601	0.691	0.562	0.675	0.639
	Disgusted	0.622	0.661	0.676	0.590	0.445	0.536
	Tame/safe	0.505	0.826	0.635	0.610	0.777	0.796
	Underwhelmed	0.786	0.809	0.827	0.673	0.717	0.554
	Curious	0.666	0.738	0.742	0.769	0.506	0.692
	Liking	0.613	0.580	0.621	0.691	0.646	0.756

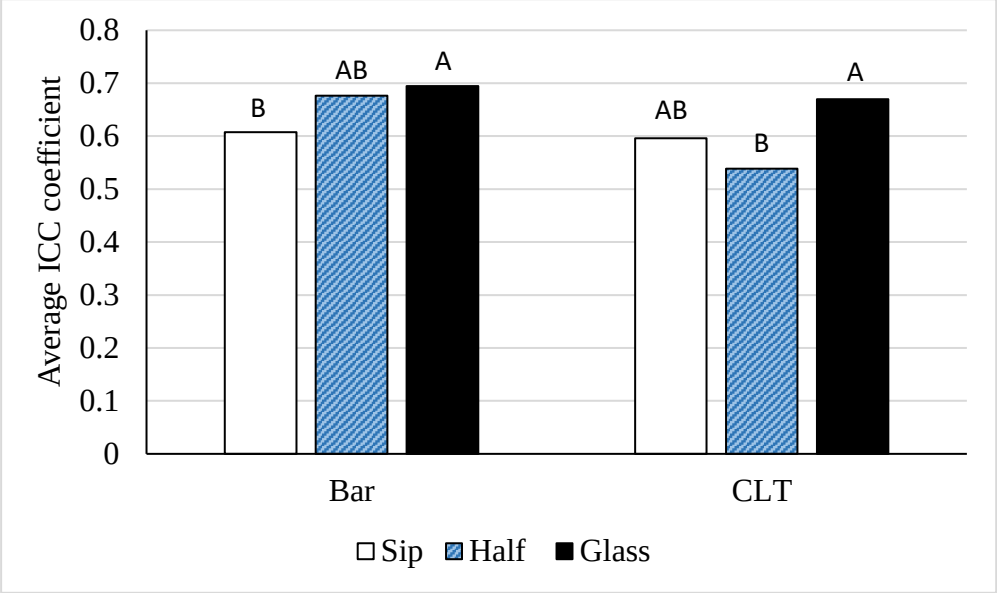


Figure 3.3: *Average ICC over all variables (10 emotion categories and liking) per consumed amount (one sip, half a glass and an entire glass) of beer and context (Bar and CLT) with Fisher's LDS post hoc groupings at $p < 0.05$ comparing amounts within each context.*

For most emotion categories and liking, higher ICC values were found for ratings given after a glass of beer than after a single sip (Table 3.4), indicating a modestly higher stability of responses collected after consumption of an entire glass of beer. This effect of consumed amount on test-retest reliability is summarised in Figure 3.3 which shows that on average, ICC's were somewhat higher after a glass of beer than when participants consumed a lower amount (F (2) = 3.1, p = 0.053).

In the CLT context there appeared to be a drop in reliability after consumption of half the glass of beer (Figure 3.3), which was observed for emotion categories Bored, Content, Excited, Disgusted and Curious (Table 3.4).

3.3.3 Effects of training on response during subsequent exposure to beer

Emotional response and liking scores were compared across sessions for participants that completed five sessions in total (n = 62). Average scores per session and F and p-values from the one-factor ANOVA comparing the sessions (Table 3.5) revealed that only two emotion categories (Content and Excited) were affected by repeated exposure. Post Hoc groupings showed that both Content and Excited were scored relatively high during the first session (the training session) and scores did not differ among subsequent (test and retest) sessions (Table 3.5).

***Table 3.5**: Mean scores (averaged over 62 participants, two context conditions and three consumed amounts) and standard errors (SE) for emotional response and liking collected during the first (training), second and third (test), fourth and fifth (retest) session with 62 participants that completed the training, test and retest sessions. All data was collected on a continuous line scale from 0 to 10. A and B indicate Fisher's LSD post hoc groupings at $p < 0.05$.*

		Session Means (n = 62)										ANOVA	
		1st		*2nd*		*3rd*		*4th*		*5th*		Session	
		M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*F*	*p*
Emotional response	Shocked	1.57	0.12	1.84	0.13	1.79	0.12	1.81	0.12	1.75	0.13	0.71	0.59
	Bored	2.97	0.17	3.23	0.16	2.89	0.16	2.67	0.14	2.79	0.16	1.81	0.12
	Content	6.81 A	0.12	6.30 B	0.13	6.24 B	0.13	6.29 B	0.13	6.11 B	0.13	4.45	**0.00**
	Excited	5.92 A	0.14	5.40 B	0.15	5.63 AB	0.14	5.67 AB	0.14	5.27 B	0.15	3.05	**0.02**
	Nostalgic	4.10	0.20	4.08	0.17	4.32	0.18	4.00	0.17	4.12	0.18	0.43	0.79
	Disconfirmed	2.04	0.15	2.22	0.15	2.13	0.14	2.06	0.14	2.06	0.16	0.25	0.91
	Disgusted	1.65	0.14	1.89	0.14	2.06	0.15	1.88	0.13	1.78	0.13	1.27	0.28
	Tame/Safe	6.08	0.15	6.00	0.15	5.80	0.16	5.76	0.17	5.82	0.18	0.71	0.58
	Underwhelmed	2.84	0.17	2.80	0.16	2.62	0.17	2.38	0.16	2.40	0.16	1.72	0.14
	Curious	4.60	0.17	4.06	0.17	4.21	0.18	4.16	0.17	4.37	0.19	1.45	0.22
	Liking	5.89	0.14	5.69	0.15	6.03	0.14	6.12	0.13	6.07	0.16	1.43	0.22

3.4. Discussion

The aim of this study was to explore the reliability of emotional response to beer from the first sip to finishing an entire glass of beer collected in a Bar environment and a CLT.

3.4.1 Context and amount impact emotional response but not liking

The results confirmed that consumption context impacts emotional response, and does so more than it impacts liking (Nijman et al., 2019). Overall higher emotion ratings were found in the Bar than in the CLT context. However, different devices were used to collect self-report data across the two context conditions. Therefore, it is unclear whether the collection of emotional response data with a smartphone in the Bar versus a desktop computer in the CLT impacted the results. Further research is needed to determine the effect of device on emotional responses.

Only two out of ten emotion categories were impacted by the consumed volume of beer. Participants felt more Content after drinking a glass of beer and more Curious after single sip. Repeated exposure to the same product has been found to decrease curiosity in previous research (Köster et al., 2003). The current study demonstrates that this reduction in curiosity can also be found during the consumption of a beer on a single occasion. However, this finding could be impacted by the fact that the beer was presented blind and thus participants might have felt curious about which beer they were drinking. This finding indicates that in blind consumer testing, first sip responses might not be completely representative of how participants feel about a beer after consuming a more representative amount.

Whilst emotional response to beer was found to vary between the Bar and CLT contexts, the context did not impact responses over the consumed amount of beer. Hence, the change in emotional response between a single sip and an entire glass of beer did not appear to be affected by context.

Liking was stable across the two contexts and between a single sip, half glass and entire glass of beer. These findings are in line with earlier research that found no differences between first and last liking scores from a temporal liking procedure for beer samples that yielded dynamic emotional responses over time and number of sips (Silva et al., 2019). It should be noted that in the current study the liking scores might have been affected by the experimental design, because liking was asked after emotional response questions in the questionnaire (King et al., 2013).

3.4.2 Similar test-retest reliability between Bar and CLT

The test-retest reliability coefficients ranged between +0.36 and +0.79 for a single sip and +0.53 and +0.84 for an entire glass of beer (Table 3.4), which was comparable to earlier findings on reliability of emotional response to food consumption by Cardello et al. (2012) who found test-retest correlations ranged between +0.37 and +0.70. In the current study, relatively low reliability coefficients for some of the emotion categories were found for measurement after consumption of half a glass of beer in the CLT context. Especially for emotion categories Content (+0.18) and Excited (+0.23) which also seemed to differ more between contexts. It is not clear what might have caused this lower reliability for half a glass of beer, nor why this was only evident in the CLT context, although it could be a result of the overall variable nature of emotion (Feldman Barrett, 2016). Considering the high number of factors that were not

controlled in the bar environment, such as music, room temperature and the presence of other customers, it is surprising to see that test-retest reliability in the bar was similar to, if not higher than, in the CLT context. This result showed that testing in a controlled environment may not necessarily provide more stable emotional response data compared to testing in a real life consumption context.

3.4.3 Higher test-retest reliability after a glass than a single sip of beer

The test-retest reliability of emotional response was found to be somewhat higher for results collected after consumption of an entire glass of beer than those collected after a single sip. This result implies that allowing participants to consume a representative volume of beer led to more stable emotion measurement, regardless of the test setting.

3.4.4 Training reduced first-order session effects on emotional response

For liking data the use of test-retest reliability has been criticised, since hedonic measurements are likely to change over time. Köster et al. (2003) have argued that repeatability is not a good criteria to measure the effectiveness of a method since first preferences seem to be a poor indication of later food choices. Since liking can change as a result of repeated exposure, it can be questioned whether emotional responses change over time and with repeated number of exposures. The third objective was to explore the effects of including a training/dummy session on the emotional response to subsequent repeated exposure to beer. Repeated exposure to the same product has been found to increase boredom and decrease curiosity (Köster et al., 2003). In the current study, repeated exposure did not affect self-reported scores on emotion categories Bored and Curious. However, scores on Content and Excited

decreased after the first training session and remained stable over subsequent sessions, indicating that single measurements were not in line with how these emotions were reported after more than one exposure to a beer sample. This finding agrees with data from Nijman et al. (2019) who found a relatively positive emotional response during the first session when the same participants evaluated beers during three separate sessions. The findings of the current study confirmed that this effect only occurred during a first session and thus the training session appeared to counteract a first-order session effect. Therefore, a training session prior to measuring emotional response could reduce this effect for studies in which participants are asked to report their emotional response in several sessions.

It should be noted that the repetitiveness of the questionnaire and study design might have had an effect on the stability of the responses. The nature of the task can have an effect on ecological validity (Jaeger & Porcherot, 2017) and emotional response may be impacted by self-report measurement (Kaneko et al., 2018). In the current study participants might have experienced fatigue from repeatedly completing the same questionnaire, despite the use of a reduced emotion lexicon in an attempt to minimise the required time and effort. Therefore, further research is required to determine the effect of repeated exposure on product differentiation using emotional response measures, in order to ascertain that reliability does not come at the cost of detecting product differences.

3.5. Conclusion

The current study investigated several issues regarding the reliability of emotional response and liking for beer. Context was found to impact emotional response overall, but not the change in response over the course of drinking a beer from the first sip to an entire glass. Test-retest reliability was comparable between a real life environment (Bar) and controlled setting (CLT), but higher for emotion measurements taken after a glass of beer than after a single sip. Emotional responses from the first (training) session were found to differ slightly from responses in subsequent test sessions, indicating that a training or dummy session could be considered prior to measuring emotional responses to better represent responses after more than one exposure.

Although the effects in this study were modest, they support the recommendation to include more realistic contexts when measuring emotional response in sensory consumer studies, in terms of test environment but also the amount of beer that is presented and evaluated in sensory consumer studies. However, further research is needed to determine the effects of consumed amount and repeated exposure on product differentiation.

Chapter 4

4. Comparing smart-phones to desktop computers as data collection device using line scales with consumers: does the device matter?

Preliminary thoughts Chapter 4:

For the experiments described in Chapter 2 and Chapter 3, the data in the natural consumption environment was collected on participants' smartphones, whilst in the controlled settings of the sensory laboratory and central location test desktop computers were used by participants to complete the questionnaires. Consequently, there was a need to investigate whether any context effects were not simply due to the use of different devices. Henceforth, the small-scale study reported in the current chapter was conducted in order to compare self-reported emotional response data collected with a smartphone to the same data collected with a desktop computer. Furthermore, this study was taken as an opportunity to confirm findings in Chapter 2 regarding the effect of consumed quantity of beer on emotional response and to determine how participants perceived the taste, flavour and mouthfeel of the beers that would be used in the final experiment on the use of Virtual Reality to simulate context in a controlled setting.

Comparing smart-phones to desktop computers as data collection device using line scales with consumers: does the device matter?

Marit Nijman[1], Jared Chua[1], Rebecca Ford[1]

[1]*Sensory Science Centre, Division of Food, Nutrition and Dietetics, University of Nottingham, Sutton Bonington Campus, United Kingdom*

Key words: Device, Mobile data collection, Online questionnaires, Line scales, Emotional response, Liking, Beer, Consumed amount

Highlights

- Stronger product differentiation was found when emotional response data were collected with desktop computers than with smartphones.
- Liking scores were not affected by the device used to collect data.
- Participants felt more curious after one sip than after consuming approximately 200 ml, regardless of device.

4.0 Abstract

Smartphones can be used to collect consumer responses in real life consumption situations, but it is not clear whether this might lead to different responses compared to when computers are used, as both have a different screen size and method to enter data (touchscreens versus mouse). This study compared consumers' responses to two beers collected with participants' personal smartphones to responses collected with desktop computers. Furthermore, responses were collected after a single sip and after consuming approximately 200 ml of beer to determine the effect of consumed amount on emotional response. Regular beer consumers (N = 44) evaluated two beers, after one sip and after consuming ≈200 ml, on ten emotion categories and liking on line scales anchored at 5% and 95%, on their personal smartphone and a desktop computer in two separate sessions. Beer presentation order and device conditions were balanced over participants. Participants felt more Curious after one sip than after consuming 200 ml (F (2, 1, 1) = 14.42, p < 0.001), confirming that single sip procedures may lead to different responses compared to when more beer is consumed. Device had no overall effect on how emotional responses were scored on line scales (p > 0.1 for all variables). However, when considering the data collected with each device separately, the beers were differentiated on scores for Excited (F (1) = 4.40, p = 0.037), Nostalgic (F (1) = 5.16, p = 0.024) and Underwhelmed (F (1) = 3.97, p = 0.048) with the computer, whilst none of the emotions nor liking differed across the two beers (p > 0.10) with smartphones. It was therefore concluded that desktop computers were more sensitive than smartphones in detecting product differentiation on emotional response measured with line scales.

4.1. Introduction

Online questionnaires facilitate data collection with portable devices such as smartphones or tablets in any location. Smartphones are convenient to use when sensory consumer research takes place in natural consumption environments such as restaurants, cafes or in home use tests. Evaluating products in an environment in which the product is usually consumed is believed to increase ecological validity of sensory consumer tests (Galiñanes Plaza et al., 2019). Furthermore, hedonic ratings are known to differ between controlled settings and natural consumption environments (King et al., 2007; Meiselman et al., 2000) and there is evidence that consumption environment can impact emotional response as well (Sinesio et al., 2019).

Self-reported emotional response has been shown to differentiate between products beyond liking (Eaton et al., 2019; Ng et al., 2013a). Most verbal self-report methods involve a predetermined list of emotions, of which participants are asked to report the extent to which they experience each emotion via check-all-that-apply (CATA) questions (Chaya et al., 2015; Ng et al., 2013a), category scales (King et al., 2010) or line scales (Dorado, Chaya et al., 2016). The latter is the most sensitive in revealing differences between products (Dorado, Pérez-Hugalde et al., 2016). Consumers' emotional responses to beer, measured on line scales have been shown to vary between a sensory laboratory and a bar (Nijman et al., 2019), however, different devices were used to collect responses, namely smartphones in the bar and desktop computers in the sensory lab. Therefore it could be possible that scale use was affected by using different devices to collect emotional response. Previous research has shown that screen size can impact sensory test results with trained assessors (Fisher,

King, Castura, & Findlay, 2016) and thus the size of the screen could have an impact on consumer responses. In addition, consumers interact with the scale differently depending on device. A mouse is typically used to give scores on a line scale with a PC computer, whereas smartphones rely on a touch screen on which participants give scores using their the fingers. The current study will explore if the use of different devices to collect self-report emotional response through line scales has an impact on sensory consumer test results.

Previous research found differences in consumer response after a single sip and after consumption of a larger amount of beer (Chapter 3). The finding that emotional responses to beer were affected by context made by Nijman et al. (2019) relied on consumers' evaluation after only a single sip of beer. A single sip does not represent natural consumption and therefore may have biased the results. Silva et al. (2019) demonstrated that both the perception of sensory properties and the emotional associations for beer can change over time and number of sips. Therefore, emotional response in the current study will be measured after one sip and after the consumption of approximately 200 ml of beer, in order to investigate the effects of device on emotional response and liking scores measured after a single sip or when more representative quantity is consumed.

The objective of the current study was to determine if using a smartphone or a computer (PC) affects consumers' use of line scales to rate liking and emotional response to beer after one sip or after consuming a more representative quantity of beer (approximately 200 ml).

4.2. Material and methods

In two separate sessions participants used either a smartphone (Phone condition) or a desktop computer (PC condition) to complete a questionnaire on emotional response and liking for two beers. The experiment was approved by the Faculty of Medicine & Health Sciences Research Ethics Committee of the University of Nottingham (Ethics reference number 111-1809).

4.2.1 Participants

Participants (N = 44) were recruited among staff and student volunteers at the University of Nottingham in addition to one volunteer external to the university. Among the total of 44 participants 32 (73%) were native English speakers. The 12 (27%) non-native English speakers had been living in the UK for at least one year and were originally from Europe (n = 11) and Asia (n = 1). Participants' ages ranged from 21 to 71 with an average age of 30 years (SD = 11) and 25 (57%) were female. Participants drank beer at least once every two months and were daily users of smartphones. Pregnant women and individuals that had any reason to refrain from drinking alcoholic beverages (including declared health, religion or addiction) were excluded from participation. Participants received an inconvenience fee in recompense for their time after they had completed both sessions of the experiment.

4.2.2 Beers

An ale and lager style beer containing (3.5% alcohol by volume) were chosen as sensory stimuli. The lager was a commercial beer commonly available in the UK. The ale was produced at the Anheuser-Busch InBev pilot brewery at The University of Nottingham. The products were selected to differ in taste and

flavour, which was confirmed with a check-all-that-apply procedure completed by consumers (see section 4.2.6.2). Both beers were bottled at the Anheuser-Busch InBev pilot brewery at The University of Nottingham, and stored at 4°C until time of testing. Beers were presented blind in 330 ml amber glass beer bottles labelled with a three-digit code and in randomised order that was balanced over participants and sessions. Participants were instructed to drink straight from the bottle.

4.2.3 Experimental design

All participants completed two test sessions, one for the PC condition and one for the Phone condition at least one week apart. To balance the order of the two conditions 22 participants were allocated to the PC condition in their first session and 22 participants started with the Phone condition. All sessions took place on Tuesdays and Thursdays in 30-minute timeslots between 2:30pm and 4:30pm. As much as individual schedules allowed, participants' sessions were allocated such that each participant came in at the same time and on the same day of the week.

4.2.4 Device conditions

All sessions for the Device study took place in the sensory booths at the Sensory Science Centre of the University of Nottingham (ISO: 8589:2007). Under the PC condition participants used the desktop computer with monitor display (Asus VT168N 15.6" 10 Point Touch Monitor, Asus Technology Holland B.V., Netherlands) provided in the sensory booths and used a mouse to complete the questionnaire. For the Phone condition, participants utilised their own personal smartphones and completed the questionnaire using their finger via the touchscreen. During the Phone session participants accessed the

questionnaire through a link in an email sent to them on the morning of their session, using their personal smartphone. Four participants that were not able to open the questionnaire on their smartphone were provided with a tablet device (9.7" Apple iPad Air A1474, Apple Inc., US) to use instead.

4.2.5 Test procedure

During each session participants received instructions on the questionnaire and test protocol before entering the sensory booths. In both Device conditions the beers were presented monadically under blind conditions according to a randomised balanced design. For each beer emotional response and liking were measured twice; once after the first sip of beer and again after the consumption of a specified amount (approximately 200 ml) that was indicated by a sticker on the beer bottle. Participants were instructed to drink until the remainder of the beer in the bottle had reached the sticker. A one-minute break was enforced between the two beer samples and consumers were asked to cleanse their palate with water.

4.2.6 Questionnaires

4.2.6.1 Emotional Response and liking

Data was collected using an online questionnaire via Compusense Cloud (Version 8.0.6288.23054, Compusense Inc., Guelph, Ontario, Canada). Participant's emotional response after the first sip and after 200 ml of beer was evaluated using a previously developed beer specific emotion lexicon (Dorado, Chaya, et al., 2016; Eaton et al., 2019) consisting of ten emotion categories (Table 4.1). Each of the ten emotion categories was presented together with the associated terms and participants were asked to indicate the extent to which they were experiencing the emotions associated with those descriptors by

giving an intensity rating on continuous line scale anchored from ‘very low’ to ‘very high’ at 5% and 95% of the scale. The order of emotion categories was randomised between participants. After evaluating emotional response, participants were asked to rate overall liking on a continuous line scale, anchored at 5% and 95% with ‘dislike extremely’ and ‘like extremely’.

Table 4.1: *Emotion categories and associated terms*

Shocked	Shocked, alarmed, cheated, confused, overwhelmed, strange, weird
Bored	Bored
Content	Content, calm, comfortable, comforted, enjoyment, good, happy, nice, pleasant, pleased, relaxed, satisfied
Excited	Excited, enthusiastic, fulfilled, fun, impressed, interested, optimistic, pleasantly surprised, want, warm
Nostalgic	Nostalgic, delirious, relieved
Disconfirmed	Disappointed, dissatisfied, unpleasantly surprised
Disgusted	Disgusted, horrible, repulsed, repelled, unpleasant
Tame/safe	Tame, safe
Underwhelmed	Underwhelmed
Curious	Curious

4.2.6.2 CATA for sensory description

The two beers were designed to be perceivably different in terms of their sensory properties in order to maximise elicitation of different emotional responses related to participants’ preferences. In order to check if participants perceived the differences between the two beers, a Check-All-That-Apply (CATA) list of consumer-generated sensory descriptors was added to the questionnaire at the end of the Phone session. CATA descriptors were generated during a pre-test where naïve panel of six consumers that did not participate in any of the other sessions of the Device Study or VR study. These six untrained volunteers tasted both beers and wrote down as may descriptors as they could to describe each beer, followed by descriptors for the differences between the two beers. Each volunteer was allowed to refer to a beer flavour

wheel (Meilgaard, Dalgliesh, & Clapperton, 1979) as a tool to help them describe their perceptions. After the individual attribute generation task, the resulting terms were discussed and compared among the six volunteers. Descriptors that were perceived to elicit the same sensation, such as ‘fizzy’ and ‘carbonated’ were grouped together. The final CATA list comprised sixteen taste descriptors that each consisted of 1-3 grouped attributes. The CATA list was presented as part of the questionnaire after participants completed the emotion questionnaire for both beers during the Phone session. Participants were instructed to take one more sip of each beer before selecting the sensory attributes that they felt complied with the statement ‘Compared to the other beer, beer [3-digit code] has a ... Please select terms from the list below’. This was completed for both beers.

4.2.7 Data analysis

Statistical analyses were performed using Microsoft Excel 2013/XLSTAT (XLSTAT Version 18.07.39020, Addinsoft, NewYork, USA). The CATA data was analysed by McNemar test on the frequency with which each sensory descriptor was selected for each beer. To determine the effects of device (PC and Phone), beer (Ale and Lager) and consumed amount (one sip and 200 ml), a three-factor mixed model Analysis of Variance (ANOVA) with device, beer and consumed amount (one sip or 200 ml) as fixed factors was performed for each emotion category and liking. Mean intensity scores and standard errors (SE) for all emotional categories and liking were calculated for each beer and each Device condition. As the aim of this study was to determine the consequences of using either a smartphone or computer to collect emotional responses and liking, the two device conditions were also considered

separately to demonstrate the impact on product differentiation. The data was split by Device and a one-way ANOVA with beer as fixed factor was performed on liking and emotional response collected with the phone and PC separately.

4.3. Results

4.3.1. CATA on perceived sensory properties

Figure 4.1 displays the number of times participants selected each of the consumer-generated sensory descriptors for each of the two beers. McNemar test revealed that for the Ale the attributes Hops/floral flavour (Q (1, n = 44) = 9.63, p = 0.002), Strong overall flavour (Q (1, n = 44) = 15.75, p < 0.001), Bitter taste (Q (1, n = 44) = 4.32, p = 0.038), Strong aftertaste (Q (1, n = 44) = 9.33, p = 0.002), Citrus fruit flavour (Q (1, n = 44) = 4.05, p = 0.044) and Sharp pungent mouthfeel (Q (1, n = 44) = 4.92, p = 0.027) were selected more frequently for the Ale than the Lager. The attributes Smooth mouthfeel (Q (1, n = 44) = 3.12, p = 0.078), Light/mild/watery flavour (Q (1, n = 44) = 3.59, p < 0.001) and Flat (not fizzy) mouthfeel (Q (1, n = 44) = 8.47, p < 0.004) were selected significantly more frequently for the Lager than for the Ale. These results confirmed that the production method used to create the Ale had resulted in perceivably different sensory characteristics compared to the Lager.

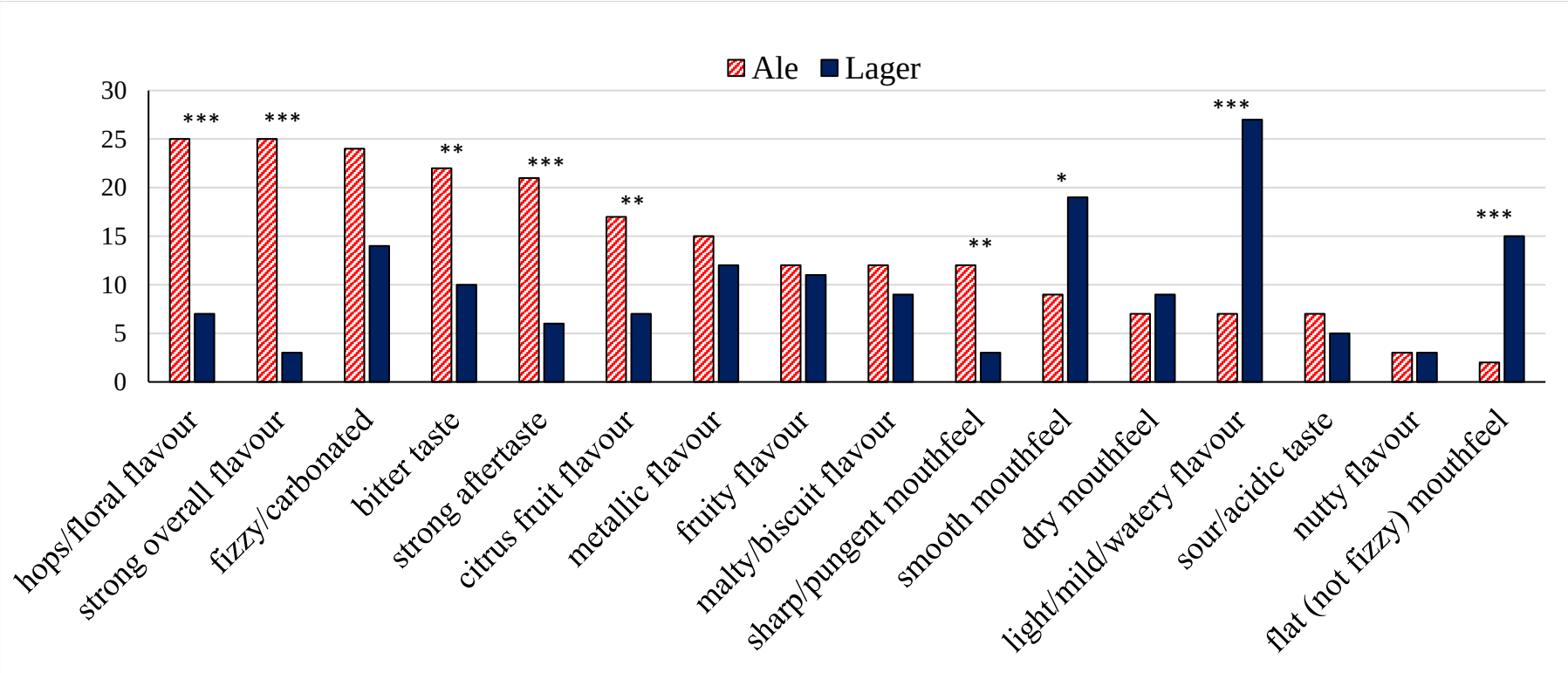


***Figure 4.1**: Number of participants out of 44 that selected attributes from the CATA list for Ale and Lager beer. *** indicates p-values < 0.01, ** <0.05 and * <0.1 from McNemar test.*

4.3.2 Emotional response and liking

As can be seen from the *F* and *p*-values in Table 4.2, the results showed no main effect of device on emotional categories, nor liking ($p > 0.1$), indicating that scores were generally similar across the two devices. The two beers were differentiated on emotion categories Excited ($p = 0.022$), Disconfirmed ($p = 0.098$), Underwhelmed ($p = 0.034$) and Curious ($p = 0.041$), but not on Liking ($p = 0.231$) (Table 4.2). The mean scores displayed in Figure 4.2 revealed that participants reported higher Excitement and Curiosity and lower scores on Disconfirmed and Underwhelmed in response to the Ale than the Lager. The amount of beer consumed only affected Curious ($p < 0.001$, Table 4.2), where mean scores were higher after one sip (*Mean* = 4.58, *SD* = 0.17) than after consuming 200 ml of beer (*Mean* = 3.68, *SE* = 0.16). The ANOVA revealed no interaction effects between device and beer, device and amount, nor beer and amount ($p > 0.10$, Table 4.2).

4.3.3 Product differentiation on emotional response and liking per device

The data collected with the Phone and PC were analysed separately to further examine the ability of each device to detect product differentiation. The mean scores per beer for all emotion categories and liking collected with each device are shown in Figure 4.3a (PC condition) and Figure 4.3b (Phone condition). When using the PC, participants reported lower scores on Underwhelmed ($F(1) = 3.97$, $p = 0.048$) and higher scores on Excited ($F(1) = 4.40$, $p = 0.037$) and Nostalgic ($F(1) = 5.16$, $p = 0.024$) for the Ale compared to the Lager. When using a Phone however, no differences between the two beers were found ($p > 0.1$). No difference in liking was found between the two beers with PC ($F(1) = 0.77$, $p = 0.38$) or Phone ($F(1) = 0.68$, $p = 0.41$).

Table 4.2**: F and p-values from three-factor ANOVA with device (PC and Phone), beer (Ale and Lager) and consumed amount (1 sip and 200 ml) as fixed factors on emotional response and liking collected from 44 participants. p-value < 0.1) are indicated in **bold

	Device (DF=1)		Beer (DF=1)		Amount (DF=1)		Device * Beer (DF=1)		Device * Amount (DF=1)		Beer * Amount (DF=1)	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Shocked	0.04	0.841	0.63	0.430	0.31	0.579	0.25	0.621	0.26	0.610	0.05	0.826
Bored	0.27	0.605	1.61	0.205	1.77	0.184	0.06	0.808	2.52	0.113	0.47	0.491
Content	0.53	0.467	1.07	0.301	0.12	0.727	0.02	0.894	0.25	0.615	0.02	0.878
Excited	0.01	0.936	5.31	**0.022**	0.38	0.536	0.44	0.508	0.14	0.708	0.06	0.809
Nostalgic	0.09	0.765	2.51	0.114	0.16	0.688	2.71	0.100	0.11	0.741	0.14	0.708
Disconfirmed	0.07	0.789	2.76	**0.098**	0.18	0.668	0.09	0.770	0.79	0.373	0.61	0.436
Disgusted	0.11	0.743	0.06	0.802	0.02	0.878	0.00	0.986	0.10	0.754	0.23	0.632
Tame/safe	1.35	0.246	1.69	0.194	0.23	0.629	0.01	0.915	2.25	0.134	0.09	0.770
Underwhelmed	0.92	0.339	4.53	**0.034**	0.26	0.610	0.47	0.494	0.03	0.863	0.66	0.418
Curious	0.20	0.654	4.22	**0.041**	14.42	**<0.001**	0.02	0.890	0.11	0.745	0.41	0.523
Liking	0.82	0.365	1.44	0.231	0.01	0.931	0.00	0.952	0.00	0.984	0.00	0.976

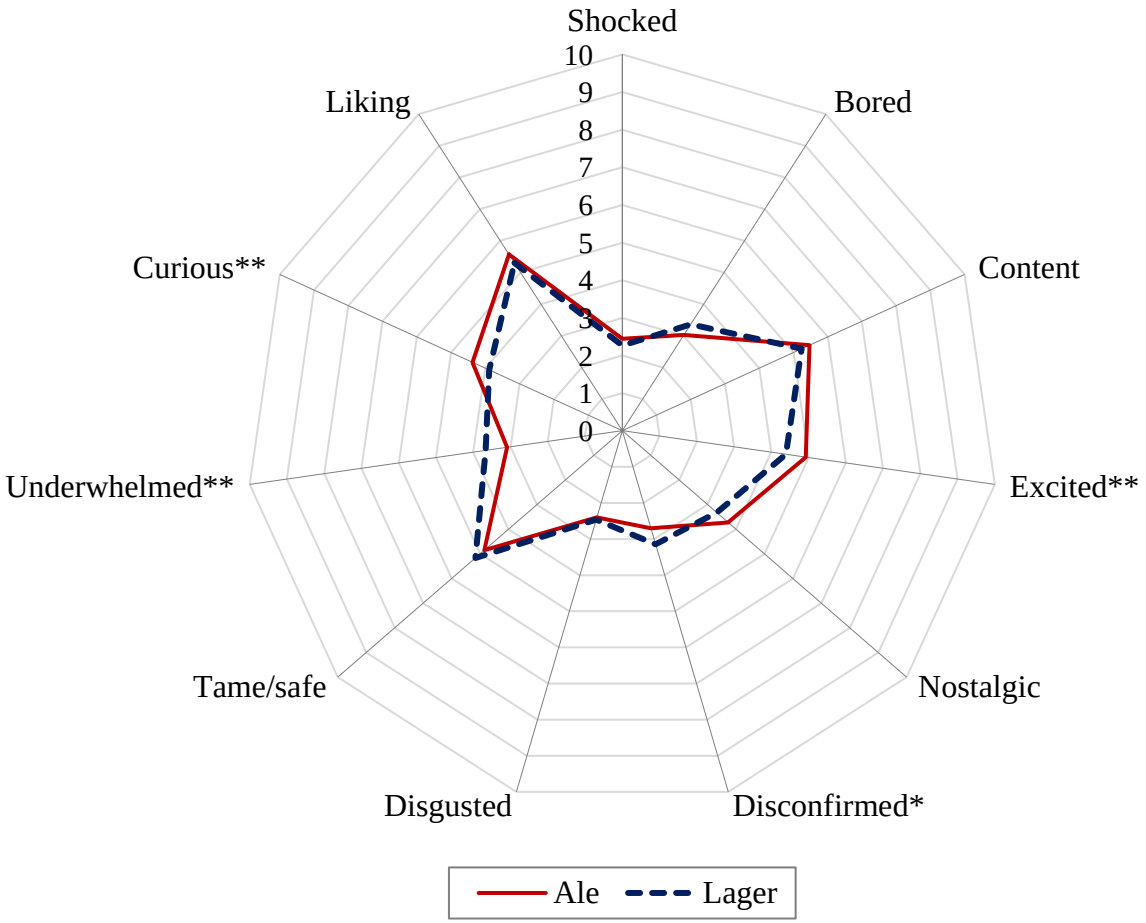


***Figure 4.2**: Mean scores over 44 participants for Lager and Ale on ten emotion categories and liking. ** indicates differences between beers at $p < 0.05$, and * at $p < 0.10$*

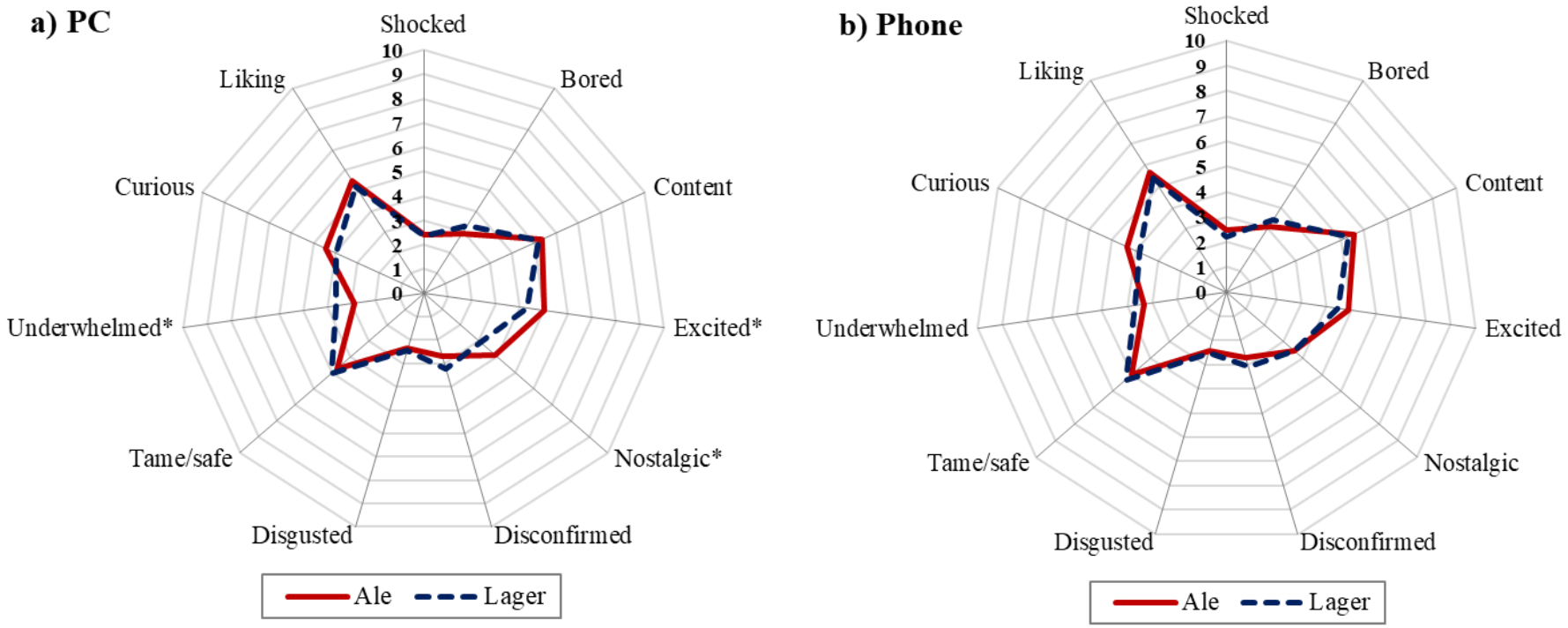


***Figure 4.3**: Mean scores for Lager and Ale on ten emotion categories and liking when evaluated by 44 participants using **a)** a PC and **b)** a Smartphone. Terms with * are different between the two beers at $p < 0.05$.*

4.4. Discussion

No overall effects of device on liking and emotional response were found (Table 4.2). This finding indicates that participants did not use a different part of the scale depending on the device they used to report their emotional response. This finding is somewhat surprising considering that previous research showed that scores collected with a laptop with monitor display were systematically lower than scores collected on an iPod or iPad (Fisher et al., 2016). However, the study by Fisher et al. (2016) investigated sensory descriptive data form trained panellists, not consumer responses. No previous research has been published on the effect of device on consumers' emotional response ratings.

The result that participants felt more Curious after one sip of beer than after consuming approximately 200 ml confirms the finding in previous research (see Chapter 3) that the amount of beer that is consumed prior to evaluation of emotional response can have an effect on the ratings for some emotion categories in the lexicon that was used in this study.

This study found no interaction effects between device and product (Table 4.2), indicating that product rank order on the emotion categories and liking did not depend on the device that was used to capture the responses. However, when considering the results separately per device, it is clear that although the response profiles looked similar across the two devices, data collection with a desktop computer resulted in a stronger product differentiation on emotional response than when participants used a smartphone to complete the questionnaire (Figure 4.3). Again this result differed from findings by Fisher et

al. (2016), who concluded that product discrimination across iPod, iPad and laptop with monitor display was similar. The limited number of participants in this study should be acknowledged as a limitation, and thus more research with larger numbers of participants would be beneficial to confirm or disprove these findings.

It is not clear whether the decrease in sensitivity when using smartphones in the current study is due to the smaller screen size compared to a monitor display. Another possible explanation could be the fact that participants used a mouse rather than their finger to mark the line scales. In the study by fisher (Fisher et al., 2016) the trained panellists used a stylus to mark the attribute scales which might have resulted in more accurate scores. A finger covers a larger surface area of the line scale on a small screen of a smartphone, possibly making the ratings less accurate than when a mouse was used on a larger screen with the desktop computer. More research is needed to determine whether asking participants to use a stylus instead of their finger would increase sensitivity. Furthermore, it was not clear if participants were using their phone in landscape or portrait view when doing the questionnaire and whether this could make a difference.

Nijman et al. (2019) found stronger product differentiation between an Ale and Lager beer when self-reported emotional responses were collected with smartphones in a bar environment than on desktop computers in a sensory lab. In the current study, collecting responses with a phone was less likely to detect differences between beers than using a desktop computer. This shows that differences in emotional response between the two beers found in the study by Nijman et al. (2019) were not due to the device that was used to collect data,

and actually suggests that using smartphones to collect data may have underestimated the differences in emotional response between the two beers when data was collected in the bar context.

4.5. Conclusion

A participants' personal smartphones can be convenient to use as the device on which sensory consumer data is collected in real life consumption situations. The current study examined the effect of using smartphones to collect self-reported emotional response data compared to using desktop computers and found that, whilst line scales presented on smartphones were less sensitive in detecting differences in emotional response between two beers than when presented on a desktop computer with monitor display, liking scores were not affected by the device used. More research is needed to determine if the effect on emotional response was due to the difference in screen size or because of using a mouse versus a finger. Researchers that want to collect emotional responses with smartphones should be aware of the consequences regarding the sensitivity to detect differences between products, and consider the risk that existing differences might not be detected.

Chapter 5

5. Simulating context with Virtual Reality: An exploration of self-reported emotional response, participant engagement and practical considerations

Preliminary thoughts Chapter 5:

The experiment described in Chapter 2 (Nijman et al., 2019) revealed that evoking a context in a sensory laboratory setting resulted in a more similar emotional response to a real bar environment with participants that were context-sensitive. This finding seemed to indicate that it was possible to improve ecological validity in a controlled setting. The experiment that is described in the current chapter builds on from this finding and explores the use of Virtual Reality as a way to bring relevant consumption context into the sensory laboratory.

In Chapter 2, first-order session effects on emotional response were found when the same participants completed several test sessions, leading to relatively positive emotional responses in the first session. In Chapter 3 a training session prior to evaluation led to more stable emotional response during subsequent test sessions. In the current chapter the efficiency of a Virtual Reality training to reduce first-order session effects will be investigated. The beers used in the current experiment were shown to differ in sensory perception in Chapter 4.

Simulating context with Virtual Reality: An exploration of self-reported emotional response, participant engagement, and practical considerations

Marit Nijman[1], Claire Humphreys[1], Qian Yang[1], Rebecca Ford[1]

[1]*Sensory Science Centre, Division of Food, Nutrition and Dietetics, University of Nottingham, Sutton Bonington Campus, United Kingdom*

Key words: Context, Virtual Reality, Evoking Context, Emotional response, Liking, Beer, Consumed amount

Highlights:

- VR led to higher levels of participant engagement than evoking context with pictures and sound recordings
- High levels of novelty related to VR impacted emotional response to beer but a training session prior to a data collection session in VR was observed to reduce novelty and associated effects.
- Tracking devices allowed participants to consume products independently while maintaining full immersion in VR.
- Audio, the time spent in VR, and realistically simulated presence of other people improved realism and immersion in VR context.

5.0 Abstract

Context can have significant impact on liking, emotional response and product choice, and Virtual Reality (VR) shows potential as a tool to evoke realistic consumption contexts in controlled consumer testing environments. This study compared an innovative approach to VR to a less immersive technique (pictures and sound recordings) to simulate a consumption environment for beer (i.e. bar), in terms of practical implications and the effects on participant engagement when measuring hedonic and emotional responses to beer. Consumers (N = 27) evaluated liking and emotional response on line scales using a beer-specific emotion lexicon for two contrasting beers (Lager, Ale) under the two context conditions. The novel VR method projected 360° video footage taken in a bar around a VR model for optimal realism and an integrated questionnaire, using a pen and beer bottles attached to tracking devices. VR training prior to data collection aimed to reduce errors associated with the novelty of the technology. Participant engagement was measured using a questionnaire and an interview. Results revealed unique insights in the impact of each method for context evocation on product differentiation, participant engagement and behaviour. VR showed clear advantages in terms of participant engagement compared to a less immersive approach (pictures and sound recordings). Audio, the time spent in VR, and realistically simulated presence of other people were identified as key elements that improved realism and immersion of the VR context, whilst low image quality and technical problems had the opposite effect. This research provides valuable information to guide decisions regarding methods to bring appropriate consumption contexts into controlled sensory test environments.

5.1. Introduction

5.1.1 Relevance of context in sensory consumer research

Context-related research has gained great interest within the field of consumer sensory science (Jaeger, Hort, et al., 2017; Jaeger & Porcherot, 2017). In standard sensory test locations a lot of effort is made to keep external sensory signals to an absolute minimum in order to ensure internal validity, which is achieved by having individual booths with neutral coloured walls and furniture in a quiet, odourless environment with temperature and humidity control (Lawless & Heymann, 1999). However, this setting does little to provide a representative consumption situation and therefore is likely to lack predictive power on how consumers experience products in real life. Hence, the measurement of consumers' responses to products in a standard sensory lab setting is likely to have low external validity. Conducting sensory consumer studies in real life consumption environments would provide higher external validity but allows for less experimental control, and thus reduces the ability to explain causal relationships between a stimulus and response. In other words, real life consumption environments offer lower internal validity than controlled settings (Galiñanes Plaza et al., 2019). Re-creating an appropriate consumption situation in a controlled environment provides consumers with relevant context while still allowing for experimental control. In theory this approach would optimise both internal and external validity. Different approaches to creating context in a controlled setting include written scenario's (Dorado, Chaya, et al., 2016; Hein et al., 2012; Piqueras-Fiszman & Jaeger, 2014a), immersive technology (Bangcuyo et al., 2015) and decorating a room to look like a bar (Sester et al., 2013) or even an aeroplane (Holthuysen et al.,

2017). Context is known to affect hedonic ratings as well as choice behaviour. Several studies have shown that depending on the product, liking scores can vary between controlled testing environments and more real life consumption environments (Edwards et al., 2003; King et al., 2007; Meiselman et al., 2000). Furthermore, beverage choices can be manipulated by changing elements in a simulated bar setting (Sester et al., 2013).

5.1.2 Virtual Reality

Virtual reality (VR) can be used to immerse individuals in a wide range of environments while maintaining experimental control (Jaeger & Porcherot, 2017). One of the limitations within context research is that findings from one context are not valid for product consumption in other contexts (Meiselman et al., 2000). Testing in one natural consumption environment does not necessarily mean that the results represent how consumers react to product 'in the real world', because of the many different contexts foods and drinks can be consumed in. Flexible methods to realistically evoke a variety of different contexts could be very useful for sensory consumer studies where it is important to explore the stability of responses over different contexts, or could be used to determine appropriate consumption situations for products under testing (Delarue et al., 2019). The Oxford English dictionary defines virtual reality as "The computer-generated simulation of a three-dimensional image or environment that can be interacted with in a seemingly real or physical way by a person using special electronic equipment, such as a helmet with a screen inside or gloves fitted with sensors." (Oxford Dictionaries, 2019, «Virtual Reality», https://en.oxforddictionaries.com/ viewed online 7th May 2019). When the term Virtual Reality is used in this paper it will refer to a system

where the individual's eyes are covered by a virtual reality headset with a screen inside. Whenever studies are discussed that used other methods than a VR headset to simulate a virtual environment, such as video screens or immersive rooms, these will be referred to as 'immersive technique' and the method will be specified accordingly. When the term 'evoked context' is used this will refer to a method used by Nijman et al. (2019) during which participants are exposure to audio and images and asked to imagine being in the depicted environment.

5.1.3 Context impacts emotional response

Context has been found to be especially relevant when measuring emotional response (Dorado, Chaya, et al., 2016; Nijman et al., 2019; Piqueras-Fiszman & Jaeger, 2015; Prescott, 2017). Emotion can be defined as an event-focussed, fast process consisting of relevance-based emotion elicitation mechanisms that shape multiple emotional responses (Coppin & Sander, 2016). Emotional response can provide additional information beyond hedonic scores (Gutjar, Dalenberg, et al., 2015; Ng et al., 2013a) and for this reason emotion measurement has gained popularity in consumer testing in recent years (Jaeger, Hort, et al., 2017; Prescott, 2017). Several studies have shown that beers can be a suitable stimulus to evoke emotional response because consumers have emotional associations with this product category (Silva et al., 2016; Silva et al., 2014; Silva et al., 2015) and beers can be differentiated by their emotional response (Chaya et al., 2015; Dorado, Chaya, et al., 2016; Eaton et al., 2019). In a recent study by Nijman et al. (2019) it was found that within a context-sensitive consumer cluster, product differentiation on emotional response for two beers was greater in a real life environment (Bar) than in a sensory

laboratory. In the same experiment evoking a context within a controlled test setting increased product differentiation on emotional response among context-sensitive consumers (Nijman et al., 2019), suggesting that evoking context can lead to more accurate predictions of consumers' emotional responses to beer. The current study investigates the use of VR to simulate a bar context for the purpose of measuring consumers' emotional responses to beer.

5.1.4 Participant engagement

Immersive techniques generally have a positive effect on participant engagement (Galiñanes Plaza et al., 2019) and have been found to provide higher levels of engagement than a sensory lab (Sinesio et al., 2019). High engagement is assumed to indicate that participants are more involved with answering the questions and thus more concentrated and motivated to give true responses. Bangcuyo et al. (2015) developed a measurement tool for engagement and used it to demonstrate a higher level of engagement among participants that evaluated coffee (Bangcuyo et al., 2015) in an immersive room with video wall, compared to when the same products were evaluated under standard laboratory settings. The same finding was made when this tool was used with participants that evaluated biscuits in an immersive room and standard sensory laboratory (Hathaway & Simons, 2017). This tool focusses on measuring eight dimensions that combined are assumed to give a measure of engagement. These eight dimensions are Usability, Environmental Aesthetics, Novelty, Involvement, Sensory Awareness, Immersion, Realism and Distraction (Bangcuyo et al., 2015). The current study measures participant engagement using these eight dimensions and aims to gain a deeper insight

into the effect of VR on novelty, immersion, realism and distraction specifically.

5.1.4.1 Novelty

VR is a new technology that most consumers are not familiar with (Andersen et al., 2019). While novelty can increase engagement (Bangcuyo et al., 2015), it might also impact emotional response, as recent evidence has shown that exposing participants to a context in VR can lead to higher levels of excitement compared to a method using a picture and the participants' imagination (Andersen et al., 2019). This raises questions on the consequences of a high level of novelty for the results obtained when using VR to evoke context. Does the exposure to VR lead to higher feelings of excitement and does this impact the intensity of overall emotion scores or product differentiation on emotional response that would not be experienced in 'real life' scenarios? And if so, does this effect reduce over time as participants get more familiar with VR and the technology becomes less novel? First order effect are a well-known phenomenon in sensory research (Wakeling & MacFie, 1995). For this reason sample presentation orders are randomized and researchers often choose to present participants with a 'dummy' sample before exposing them to the products of interest. First order effects are also known to impact emotional response when it comes to the order in which products are presented to consumers (Dorado, Pérez-Hugalde, et al., 2016). Furthermore, there is evidence that overall emotional response to beer is more positive during a first session than during sessions that follow when evaluated by the same consumers (Nijman et al., 2019). It would be of interest to know whether there are any effects on emotion data due to initial excitement related to the

novel nature of VR, and if there are, whether these effects become less apparent when a participant is exposed to the technology a second time.

5.1.4.2 Immersion

VR and other immersive techniques provide a high level of immersion which might be beneficial in evoking context in a controlled environment. Less immersive techniques such as written scenarios and exposure to pictures and sound recordings can have an effect on product differentiation with emotional response (Nijman et al., 2019; Piqueras-Fiszman & Jaeger, 2014c). Although these methods are easy to apply in a controlled setting and offer versatility in terms of the number of different contexts that can be evoked, their effectiveness also depends on a participant's imagination and willingness to adhere to the instructions of engaging with the evoked context (Jaeger & Porcherot, 2017; Köster, 2003). More immersive strategies of evoking context that demand less effort from participants might approach results from real life consumption environments more closely. Andersen et al. (2019) found that a simulated beach environment in VR had a stronger effect on the desire for hot and cold beverages than when combining participants' imagination with exposure to a picture of a beach. Furthermore, VR can yield higher self-reported experience of immersion than a real environment (Van Herpen, Van den Broek, Van Trijp, & Yu, 2016) and the immersive effect of VR has the power to change sensory perception (Stelick, Penano, Riak, & Dando, 2018). This shows that VR could be a much more powerful tool than methods that rely more on participants' imagination such as written scenarios and exposure to images. The current study will measure the immersive effect of VR

compared to evoking context with audio and images and explore the relation between immersion and realism.

5.1.4.3 Realism

Realism is an important factor for engagement (Bangcuyo, Smith et al. 2015). Virtual reality is a method to simulate reality and any simulation has limitations. The use of electronic devices might compromise the realism of participants' experience even when participants are used to the type of technology (Galiñanes Plaza et al., 2019). Not much is known about which elements contribute to making a context evocation in VR convincing to sensory test participants. However, there is evidence that in an immersive room with video wall participants pay most attention to visual aspects, followed by auditory aspects, whilst olfactory aspects are paid the least attention (Liu et al., 2019). The current study aims to explore elements in VR that increase or reduce perceived realism.

5.1.4.4 Distraction

Standard sensory booths facilitate focus on the product by minimising any distracting external factors. In contrast, real life situations typically have many noise factors present that distract attention away from food or drink items that are being consumed. When evoking a context in controlled settings noise variables like auditory and visual stimuli are generally introduced in a controlled way. Immersive technologies have been shown to induce similarly higher levels of distraction as a real pub environment compared to sensory laboratory setting (Sinesio et al., 2019). Absence of distraction improves participant engagement (Bangcuyo et al., 2015) and therefore it would be of

interest to know whether participants experience VR to be distracting while they perform their product evaluations.

5.1.5 Practical considerations

There are several practical considerations to using VR in sensory testing such as the collection of data, product consumption, providing realistic social contexts and ensuring participant comfort. The current study explored innovative solutions to these limitations.

5.1.5.1 Consuming products in VR

One of the main challenges of applying VR in sensory research is to allow participants to consume foods and drinks while wearing a VR headset. A VR headset completely cancels out participants' sight of physical reality and so the food or drink vessels cannot be seen. Possible solutions are to remove the headset for product consumption or for a researcher to assist participants by handing them the samples (Sinesio et al., 2019). However, more sophisticated solutions exist such as systems where a camera on the headset allows for real-time tracking of movements of the participant's hands and objects in the foreground. Sinesio et al. (2019) found that this technology allowed participants to consume beers without the help of a researcher, but a bit of practice was required before participants felt confident with it and some participants complained about difficulties with drinking the beer (Sinesio et al., 2019). Another solution is to model physical objects in VR with the use of tracking devices. Trackers are small motion tracking devices that can be attached to physical objects and allow a VR system to register the movements of the physical objects and model them in the virtual reality. This way participants can see objects like food and drink vessels in VR. The use of

trackers would therefore be a useful strategy to allow participants to independently consume products while maintaining immersion in VR. On the other hand, problems with the trackers could disrupt natural drinking behaviour and therefore interfere with sensory perception and bias emotional response. Since sensory testing sometimes relies on measurement after consumption of only a small amount of sample, it is of interest to know if problems are more apparent for a first sip compared to after several sips. In the current study the use of trackers and their effects on responses after one sip and the consumption of a larger amount were investigated.

5.1.5.2 Questionnaire design

Measurement of consumers' responses during context exposure is key because effects of VR fade quickly after the headset is removed (Andersen et al., 2019). This might be even more true for emotional response measurement due to the fast nature of emotions (Coppin & Sander, 2016). The use of controllers and buttons to answer questions in VR are not always considered easy to use by participants (Stelick & Dando, 2018; Stelick et al., 2018). Furthermore, controllers might be associated with computer games and therefore it could be speculated that their use reduces realism. Virtual reality offers a lot of opportunities for more interactive methods of data collection, allowing any virtual object, including the participants themselves, to be used as a questionnaire tool. For example, product preference can be registered simply by moving the product of choice or by asking participants to walk towards a virtual object. Instructions or questions do not have to be presented on a white surface or as 'floating words' but can be integrated into the virtual environment or presented verbally by a virtual character (avatar) (Dzardanova,

Kasapakis, & Gavalas, 2018). These features could potentially make questionnaire completion easier and more engaging. In the current study the possibility of capturing questionnaire answers in VR using object-tracking and a questionnaire that was integrated in virtual design objects of the VR environment was explored.

5.1.5.3 Social context in VR

If a context simulation aims to represent a consumption situation that typically takes place in the presence of other people there is a challenge to realistically evoke the social aspects of the situation in VR. Social interaction and even the mere presence of other people can impact consumption behaviour (Bell & Pliner, 2003; Cruwys, Bevelander, & Hermans, 2015; Patel & Schlundt, 2001) and possibly product liking (King et al., 2004). Lifelike footage of individuals can be captured with 360° video which enables a simulated presence of other people in VR. However, 360° video does not allow social interaction with the individuals that are visible in the video. Environments and aspects that are computer-model based, such as avatars, do allow interaction. However, these avatars do not (currently) reach the level of realism that can be obtained by filming real human beings via 360° video. Their appearance remains cartoon-like, such as characters in video games and therefore they lack realism. Even when the avatar's movement and speech is operated by an actual individual, participants tend to not consider the avatar to be a real person (Dzardanova et al., 2018). Alcoholic beverages are associated with social situations (Keesman et al., 2018) and therefore simply having people around does not account for the social aspect of 'having a beer in a bar'. Immersive rooms that allow several participants to participate in a sensory test at the same time and interact

with each other have a clear advantage over VR in terms of opportunities to facilitate social interaction. Therefore, creative approaches should be considered in dealing with the lack of social interaction in VR. In this study the use of a written scenario prior to VR exposure was explored as a way to offer an explanation for the lack of social interaction in a bar context where social interaction would normally be expected.

5.1.5.4 Comfort

Wearing VR goggles can cause motion sickness or dizziness (Sinesio et al., 2019; Stelick et al., 2018). Andersen et al. (2019) recorded usage problems with VR such as dizziness and weight of goggles and found that eight out of 60 participants (13%) experienced dizziness. Other problems that participants reported were related to wearing prescription glasses underneath the headset, pressure and weight of goggles and blurry vision (Andersen et al., 2019). As discomfort might impact emotional response and could be a cause of distraction it is considered relevant to investigate causes of discomfort related to VR.

5.1.6 Aims of the current study

This study aims to explore the effect of simulating consumption context in VR on participant engagement compared to evoking context with audio and images when measuring hedonic and emotional responses to beer. In addition practical implications of VR related to measuring responses, consuming beverages, replicating social context and participant comfort will be studied with the aim of highlighting relevant considerations that are to be made when using VR for sensory consumer research.

5.2. Material and methods

VR was used to simulate a bar environment where beer would typically be consumed (VR condition) to a less immersive method, namely a picture slideshow and sound recordings to elicit the same context (Evoked condition). This study was approved by the Faculty of Medicine & Health Sciences Research Ethics Committee of the University of Nottingham (Ethics reference number 111-1809).

5.2.1 Participants

A total of 27 participants were recruited among students (81%) and staff (15%) at the University of Nottingham in addition to one local volunteer external to the university (4%). All participants were required to complete a recruitment survey prior to participation. Thirteen participants were female (48%) and 14 male (52%). Participants' ages ranged from 20 to 67 with an average age of 28 (SD = 9.8). A majority of 23 participants (85%) were native English speakers. The non-native English speakers had been living in the UK for at least 3 years and were originally from Europe (11%) and Central America (4%). When asked if they had ever experienced VR before a majority of 59% of the participants said 'no', 37% said 'yes, once or twice', and one participant (4%) replied with 'yes multiple times' and explained their experience was with VR gaming. All participants drank beer at least once every two months and had been to the Student Union bar of the University of Nottingham at least once. Pregnant women and individuals that had any reason to refrain from drinking alcoholic beverages (including declared health, religion or addiction) were excluded from participation. Participants received an inconvenience allowance at the end of the study in recompense for their time.

5.2.2 Beers and sample presentation

An ale and lager style beer were chosen as sensory stimuli for all sessions. Both beers contained 3.5% alcohol by volume (ABV). Beers with a relatively low alcohol percentage were chosen for ethical considerations regarding the exposure of participants to alcohol, but also to ensure that the measured responses were a result of the beer drinking experience rather than the effects of alcohol. The Lager was a commercial beer commonly available in the UK. The Ale was custom made to match the Lager in alcohol percentage but produced with an ale yeast and with additional bittering hops as well as aroma hops to create a different flavour profile and stronger bitter taste compared to the Lager. In order to test the assumption that participants perceived the designed differences between the two beers, 47 beer consumers evaluated the two beers using a Check-All-That-Apply (CATA) list of consumer-generated sensory descriptors during a pre-test. The results confirmed the perceptual differences (Chapter 4). Both beers were bottled into unbranded amber 330 ml glass beer bottles at the Anheuser-Busch InBev pilot brewery at The University of Nottingham. All beers were stored at 4°C until time of testing and taken out of cold storage shortly (up to 5 minutes) before being served. In all sessions the beers were presented under blind conditions in a randomised order that was balanced over participants and sessions. The beers were served in bottles labelled with a three-digit code and participants were instructed to drink straight from the bottle. This approach was decided upon after pre-tests with the VR equipment revealed that tilting a glass enough to drink from it while wearing the VR headset was problematic. In addition, a bottle gave a lower risk of spilling than a glass due to the smaller opening.

5.2.3 Written scenario

During each session, exposure of the participant to the context condition was preceded by reading a written scenario that stated: *"Think of one of your friends that you would meet up with in a bar. Imagine that friend. Imagine having agreed to meet this friend in a bar in the late afternoon. You arrive a bit early and your friend just texted that he/she is running late, so you decide to go ahead and have a beer while you wait."*

5.2.4 Emotion questionnaire

Participant's self-reported emotional response was measured using a previously developed beer-specific emotion lexicon (Dorado, Chaya, et al., 2016; Eaton et al., 2019) consisting of ten emotion categories (Table 5.1). Each of the ten emotion categories was presented together with the associated terms and participants were asked to indicate the extent to which they were experiencing the emotions associated with those descriptors by giving an intensity rating on continuous line scales, anchored from 'very low' to 'very high' at 5% and 95% of the scale. The order of the emotion categories was randomised over participants and sessions. After evaluating emotional response, participants were asked to rate overall liking on a continuous line scale, anchored at 5% and 95% with 'dislike extremely' and 'like extremely'. For each beer, emotional response and liking were measured twice; once after the first sip and again after the consumption of approximately 200 ml. The amount of 200 ml was indicated by the location of the sticker with the 3-digit code on the bottle. Participants were instructed to drink until the remainder of the beer in the bottle had reached the top of sticker. A one-minute break was enforced between the two beer samples.

Table 5.1: *Emotion categories and associated terms*

Shocked	Shocked, alarmed, cheated, confused, overwhelmed, strange, weird
Bored	Bored
Content	Content, calm, comfortable, comforted, enjoyment, good, happy, nice, pleasant, pleased, relaxed, satisfied
Excited	Excited, enthusiastic, fulfilled, fun, impressed, interested, optimistic, pleasantly surprised, want, warm
Nostalgic	Nostalgic, delirious, relieved
Disconfirmed	Disappointed, dissatisfied, unpleasantly surprised
Disgusted	Disgusted, horrible, repulsed, repelled, unpleasant
Tame/safe	Tame, safe
Underwhelmed	Underwhelmed
Curious	Curious

5.2.5 Engagement questionnaire

Engagement was measured with an online questionnaire via Compusense Cloud (Version 8.0.6288.23054, Compusense Inc., Guelph, Ontario, Canada). The engagement questionnaire was an adaptation of the questionnaire used by Bangcuyo et al. (2015) and consisted of 20 items that were aimed to measure participants' experience with regards to specific statements comprising of their eight dimensions of engagement. Usability, environmental aesthetics, novelty, involvement and immersion were measured on a seven-point Likert scale (coded -3 to +3). Bangcuyo et al. (2015) et al used a 5-point Likert scale ranging from -2 to +2 but for the current study a more sensitive scale was used to allow for a deeper exploration of differences between the two context conditions. Sensory awareness, distraction and realism, in addition to two items related to participants' comfort based on findings from Andersen et al. (2019), and one statement related to difficulties with drinking the beer, were rated using a 7-point categorical scale ranging from 1 (not at all) to 7 (very much). Specific questions and statements comprising each dimension of the questionnaire are shown in Table 5.2.

***Table 5.2**: Engagement questionnaire. [R] indicates reverse coding.*

Dimension	Statement/Question
Usability	The testing environment assisted in my evaluations of the samples.
Environmental Aesthetics	The testing environment was appealing.
	The testing environment engaged my senses.
Novelty	The testing environment incited my curiosity.
	The testing environment distracted me. [R]
Involvement	The testing experience was boring. [R]
	The testing experience was fun.
	I was engaged in the sensory task I performed.
Sensory Awareness	How completely were all of your senses engaged by the testing environment?
	How much did the visual aspects of the testing environment involve you?
	How much did the auditory aspects of the testing environment involve you?
Immersion	I felt like I was in a bar.
	I lost track of time.
Realism	How disconnected did you feel from the testing environment? [R]
	How much did your experiences in the testing environment seem consistent with your real-world experiences?
	How completely did you feel immersed in the testing environment?
	How involved were you in the testing environment experience?
Distraction	How aware were you of events occurring in the real world around you? [R]
	How quickly did you adjust to the testing environment experience?
	How much did the testing environment interfere or distract you from performing your sensory evaluation? [R]
Comfort	Did you experience dizziness?
	Did you experience discomfort in this test environment?
Drinking Beer	Did you experience problems with consuming the beer?

5.2.6 Interview

Each participant was interviewed face-to-face by the researcher that conducted the VR training session. The researcher asked pre-determined questions covering eight themes (Table 5.3) that were based on literature and pre-tests with the VR equipment. In addition to the scripted questions, the researcher asked probing questions following the participant's answers to explore additional themes that came up, and to encourage the participant to further explain the answers. Questions from the interview script were not asked if the topic was already covered by the answers to previous questions. The objective of the interview was to gain a deeper understanding of how VR impacted participant engagement as well as how participants experienced the different practical implications of the VR elements used in this study, such as the integrated questionnaire, the tracking of objects and comfort of the headset. Two out of twenty-seven participants did not give permission to be filmed in their written consent and therefore only audio recordings were made of their interview. For the remaining 25 participants both video and audio recordings were made with a video camera (Canon EOS 6D, Canon Inc., Japan).

***Table 5.3**: Interview script*

Theme	**Interview questions**
First reaction:	How was that/what did you think? How did the experience make you feel?
Environment:	What did you think about the environment you were in? Can you describe the environment? (Did you recognize the bar?) Would this be a setting that you would encounter in your real life?
Realism:	What made the environment realistic/not realistic? What did you think about objects/furniture/sound/people/not seeing yourself
Distraction:	Would you say you felt distracted from the task by the environment? What was distracting?
Drinking beer in VR:	Were you drinking the beers as you normally would in a bar? Can you describe your experience with drinking the beer? Did you feel reluctant to drink from the VR bottle?
Questionnaire:	What did you think of the questionnaire? What did you think about the way you were answering the questions?
Social context:	Was the written scenario realistic? Would this scenario be applicable to you? What did you think about having people around you?
Comfort:	Was there anything about the VR experience that made you feel uncomfortable? Was the headset comfortable? Did you experience any dizziness or motion sickness?

5.2.7 Context conditions

5.2.7.1 The VR bar

In the VR condition participants were seated at a table in a separate test room of the Sensory Science Centre as displayed in Figure 5.1. A three-dimensional 360° representation of the Student Union Bar was presented through a VR headset (HTC Vive Pro VR Virtual Reality Headset 2018/v2, HTC Corporation, Taiwan) and run on a specialised gaming PC (Coffee Lake Intel Core i5 8400 and NVIDIA GTX 1070 Ti, Scan Computers International Ltd, UK). A disposable hygiene cover was placed inside the headset before the start of each session. The virtual environment consisted of a combined 360° video and a computer model environment that was created using the Unity game development platform (Unity Technologies https://unity.com/). The 360° video lasted 22 minutes and 23 seconds and was recorded with a 360° camera (RICOH THETA V 360 Degree camera, The Ricoh Company, Ltd., Japan). In order to guarantee participants would continuously be exposed to the VR Bar environment, the footage was looped so that the video would immediately restart after reaching the end of the recording. The camera was positioned in centre of a seating area of the bar and featured fourteen people that were either seated at tables around the camera or standing at the bar. The individuals remained in the same area of the bar for the entire duration of the video. Two lenses on either side of the 360° camera captured footage in 4K (3840 x 1920, 29.97fps) and the two perspectives were stitched together in editing to create the 360° view. Directional sound was recorded with the four microphones inside the 360° camera and consisted of conversations and background noises of the people in the video as well as music played at low volume.



Figure 5.1*: Representation of the VR condition during which the participant is wearing a VR headset and moves a tracked beer bottle while a second tracked beer bottle and marker pen are placed on the table.*

During the sessions participants were allowed to adjust the volume using the control buttons on the headset. The virtually-modelled features included a table, three chairs and a floor, as well as two beer bottles and a marker pen. The virtual table, chairs and floor were designed to match the interior of the Student Union Bar and to have similar dimensions as the physical chair and table at which the participant was seated throughout the session. The computer-modelled features could be interacted with, whereas the 360° video was displayed around the modelled environment and would always be an infinite distance away from the headset, i.e. the participant. Careful considerations were made over many pre-tests in the process of creating the design so that the modelled environment would blend with the 360° footage as seamlessly as possible. Three tracking devices (HTC Vive Object/Peripheral VR Trackers, HTC Corporation, Taiwan) were used, of which two were attached to the glass bottles in which the beer was presented as shown in

Figure 5.1. A third tracking device was attached to a marker pen that was used to complete the questionnaire. The tracking devices allowed any movements that were made with the physical beer bottles and pen to be matched by their computer-modelled counterparts in VR. The computer-modelled objects were made to resemble the physical objects, including the tracking devices they were attached to. Three-digit codes to identify the beer were entered by the researcher at the start of the session and were displayed on the virtual model of the beer bottles so they were visible to the participants (Figure 5.2).

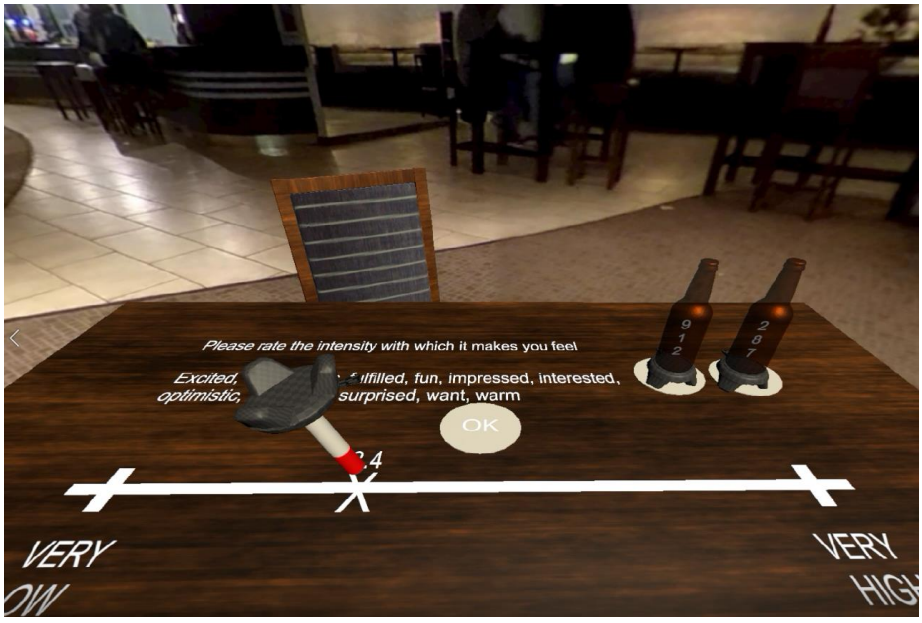


Figure 5.2*: Representation of participants' view in VR condition. Participant uses the marker pen to mark the line scale that is depicted on the table. Instructions, 'OK' beer coaster and the two beer samples are also visible on the table and around the table the 360° video of the bar is playing.*

Instructions and questionnaire questions appeared one-by-one on the surface of the virtual table and a line scale was depicted across the virtual table in front of the participant (Figure 5.2). Participants could give intensity ratings by holding the pen over the virtual line scale. The marking of the line scale was fed back to the participant by showing a number that would appear just above the scale

and that corresponded with the given score (Figure 5.2). To confirm their score, participants held the pen over a virtual beer coaster marked with ‘OK’ that was depicted on the virtual table centrally above the line scale. Confirming the score automatically triggered the next question. The flow of the questionnaire and a one-minute break between the two beer samples was controlled by the researcher who followed the progress of the questionnaire on the screen of the gaming PC. The virtual beer bottles did not show any liquid in the bottle, and so participants were not able to see how much beer they had consumed. The researcher observed the amount of beer left in the physical bottle and indicated to the participant when the remainder of beer in the bottle had reached the top of sticker by activating a ‘Stop drinking’ sign that appeared on the virtual table. Participants were instructed to drink at their usual pace until the ‘Stop drinking’ sign appeared.

5.2.7.2 The Evoked condition

Under the evoked condition each participant was provided with a tablet device (9.7” Apple iPad Air A1474, Apple Inc., US) in their sensory booth (Figure 5.3) on which they could watch a video that consisted of a picture slideshow and audio recordings taken in the same bar as the 360° video used in the VR sessions. Thirty minutes of sounds from the bar were recorded on a regular Tuesday evening, consisting of music, indistinguishable conversation and background noises and five photos were taken from different angles in the bar and were displayed as a repeating slideshow with a duration of five seconds per photo.



Figure 5.3*: Representation of Evoked condition during which participants were exposed to sound recordings and picture of the bar via a Tablet and completed the questionnaire on the PC.*

5.2.8 Procedure

5.2.8.1 Sessions

Participants tasted each of the two beers during three sessions which were at least one week apart: 1) a VR Training session, 2) a VR session and 3) an Evoked session. During the VR Training session and the VR session participants were exposed to the VR condition whilst for the Evoked session participants were exposed to the Evoked condition. The VR Training session always preceded the VR session but the order of the VR condition and Evoked condition was balanced over participants such that 14 participants were allocated to the VR condition before completing the Evoked condition and the other 13 participants were exposed to the Evoked condition before completing the VR condition. All sessions for the VR Study took place at the Sensory Science Centre of the University of Nottingham from Monday to Thursday between noon and 6pm.

5.2.8.2 VR Training

A one-on-one VR training session with a researcher was completed by each participant. The objective of the training was to familiarise participants with the VR environment, instruct them on the session protocol and allow them to practice drinking from beer bottles whilst in the VR environment, and to complete the questionnaire using the virtual pen. In addition, emotional response and liking data were collected for both beers to replicate the VR session. The training lasted one hour and consisted of verbal instructions given by the researcher, followed by three activities. The first activity was a first exposure to the VR environment during which participants practiced scoring questions using the pen and taking sips with an empty bottle and then a bottle filled with water. During this first activity participants were wearing the headset but did not cover their ears with the headphones, to allow communication between the participant and researcher. Once the participant felt comfortable they were given a short break during which the headset was taken off, the participant was given the opportunity to ask questions and asked to read the written scenario described in section 5.2.3 from a sheet of paper. For the second activity participants evaluated their emotional response and liking for the two beers with the emotion questionnaire as described in section 5.2.4 and section 5.2.7.1. During this activity the participant's ears were covered with the earphones and they no longer communicated with the researcher. The third and final activity was the interview described in section 5.2.6, which commenced immediately after participants took off the headset.

5.2.8.3 VR session

VR sessions were also conducted one-on-one with each participant and took approximately 40 minutes. At the start of the VR session the participant was reminded of the instructions given during the VR Training session. After reading the written scenario on paper the participant was assisted by the researcher to put on the VR headset. Each participant was exposed to the VR environment for one minute before the emotion questionnaire started. When the questionnaire started the participant evaluated emotional response and liking for the two beers as described in section 5.2.4 and section 5.2.7.1. Context exposure ended after the participant completed the emotion questionnaire. The researcher helped the participant remove the VR headset and the participant was presented with the engagement questionnaire described in section 5.2.5 on a tablet device (9.7" Apple iPad Air A1474, Apple Inc., US).

5.2.8.4 Evoked session

The sessions for the Evoked condition lasted approximately 30 minutes and were conducted with up to 10 participants at a time. At the start of the session participants received instructions about the test protocol before being allocated to individual sensory booths (ISO: 8589:2007). After reading the written scenario described in section 2.3 from the computer screen participants were instructed to put in a set of earphones and start the video on the tablet device. Participants were allowed to adjust the volume using the control buttons on the tablet and they were instructed to keep their earphones in and listen to the sound recordings until they had completed the emotion questionnaire (section

2.4). Participants used a desktop computer in the sensory booth to complete the emotion questionnaire and the engagement questionnaire (section 2.5).

5.2.9 Data analysis

5.2.9.1 Emotional response and Liking

One participant had to be removed from the dataset due to a sample mix-up, leaving 26 participants for analysis. Mean intensity scores and standard errors (SE) for all emotional categories and liking were calculated for each beer (Ale and Lager) from each session (VR Training, VR session and Evoked session).

To study the overall effects of context condition and its interactions with other factors, a four-factor ANOVA and Fisher's LSD post hoc test with context (VR session or Evoked session), beer (Ale or Lager), consumed amount (one sip or 200 ml) and sample position (presented first or second) was performed for each emotion category and liking.

To explore differences in response between the VR Training and VR session (the first and second time participants were exposed to VR) and the relative effects compared to the Evoked condition, a four-factor ANOVA with Session (VR Training, VR session or Evoked session), beer (Ale or Lager), consumed amount (one sip or 200 ml) and sample position (presented first or second) was performed for each emotion category and liking.

5.2.9.2 Engagement questionnaire

Principal factor analysis was performed on the rating data of the 20 items that related to the eight dimensions of engagement (Table 5.2). Cronbach's alpha was used to measure internal consistency overall and for each dimension of engagement. The ratings for each of the 20 items related to engagement were

averaged over participants. Following the procedure by Bangcuyo et al. (2015) data from relevant items for each dimension were averaged to generate a score for each dimension within each test condition (VR and Evoked). A total engagement score per test condition was generated by adding the dimension scores. The ratings for the two items related to comfort and the item on beer consumption were also averaged over participants. A mixed model two-factor ANOVA with participant as random factor and test condition as fixed factor was performed for each item, each dimension score and the total engagement score.

5.2.9.3 Interviews

All 27 interviews were transcribed verbatim using the NVivo software. Word-for-word transcription with minimal additional description of hand gestures was made of each interview based on the recordings. Hand gestures were only described when important for the understanding of spoken word. For example when participants said ‘that’ and pointed towards a specific object, the transcription included a description of object the pointing gesture was made to. For the sake of investigator triangulation, transcripts were analysed independently by two researchers following two different strategies. One researcher used the NVivo software to code the transcripts on eight pre-determined codes based on the eight themes the interview focussed on (Table 5.3), followed by a secondary round of coding to identify three additional themes for statements related to i) comparison of VR to the sensory lab, ii) Immersion and iii) Previous VR experience, resulting in a total of eleven themes. Additional subthemes were coded for Comfort (Glasses, Headset, Motion Sickness, Other), Drinking beer (Appearance of bottles, Beer itself,

Technical difficulties, (un)natural behaviour), Environment (Appropriate, Atmosphere, Recognition), Immersion (Time spent, Sound), Realism (Felt like a real pub, Being invisible, Headset, Objects, People, Sound, Time spent, Video and Model) and Social context (Interaction, People around, Written scenario). Subsequently, content within each theme was summarised. The second researcher analysed the transcripts without software and generated the key themes and subcategory lists from the transcripts which were then cross-referenced with the interview themes from Table 5.3. Subsequently a 'cut and paste' method was used in which relevant parts from the transcripts were 'cut' for each category and 'pasted' together to create a matrix and the content within each category was summarised.

The summaries were compared and discussed between the two researchers. Overall, the content of the themes and subcategories aligned between the separate analyses, although the categorisation differed slightly. Findings that both researchers agreed on were reported. Considering the relatively low number of participants, comments that were made by one participant were reported if both researchers found the comment relevant.

5.3. Results

5.3.1 Emotional response and liking

In this section the results from the emotion questionnaire completed in the two context conditions (VR session and Evoked session) will be discussed first, followed by the results from the VR Training session.

5.3.1.1 Comparing responses between the VR and Evoked context

The *F* and *p*-values from the ANOVA on context, beer, amount and sample position are shown in Table 5.4. As can be gathered from the low *p*-values (< 0.1) for context effect on four emotion categories (Bored, Content, Excited and Curious) and overall liking, responses differed between the VR session and Evoked session. When comparing the mean response scores between the VR and Evoked session (Figure 5.4 and Table 5.5) it is clear that overall, participants reported lower scores on Bored and higher scores on Content, Excited and Curious in the VR session than in the Evoked session.

When considering the *p*-values for the effect of beer in Table 5.4, differences between the two beers in terms of emotional response and liking appear limited. Under both context conditions, the extent to which participants felt Bored was higher for the Lager than the Ale (Figure 5.5a and b).

The amount of beer participants consumed before rating their emotions only affected the scores for Bored (Table 5.4). Overall, participants gave higher rating on Bored after they had consumed 200 ml (Mean = 3.1, *SE* = 0.24) than after a single sip of beer (Mean = 2.4, *SE* = 0.20).

Whether a beer sample was presented first or second affected the responses on Shocked, Content, Disconfirmed, Disgusted, Tame/safe, as well as liking.

Furthermore, the sample position effects were depended on the type of beer (Lager or Ale), and if that beer was consumed during the VR or Evoked session, which is clear from the two-way interaction between beer and sample position and the three-factor interaction between context, beer and sample position (Table 5.4). Studying three-factor interactions for sample order was not in the scope of this study. However, it can be mentioned that in the Evoked session, the Ale appeared to yield a more positive emotional response when it was presented as the second sample than when it was presented first, whereas the Lager gave a more positive emotional response when it was presented as the first sample. In the VR Training and VR session both beers appeared to yield a somewhat more positive response when presented first.

Nostalgic was the only emotion category that seemed affected by the interaction between context and beer ($p = 0.092$, Table 5.4). Under the evoked condition, the Ale was scored higher on Nostalgic than the lager, whilst this effect was not seen during the VR session (Figure 5.5a and b). Although visually, the responses appeared to differ slightly more between the Ale and the Lager during the Evoked session (Figure 5.5b) than during the VR session (Figure 5.5a), no reliable analyses could be done to further explore product differentiation within each context due to the strong effects of sample order and the low number of participants.

Other than the slight interaction between context and beer on Nostalgic and the interactions between beer*sample position, and context*beer*sample position, no interactions yielded p-values < 0.10.

Table 5.4: F and p-values from a four-factor ANOVA with context (VR and Evoked), beer (Ale and Lager), consumed amount (one sip and 200 ml) and sample position (first or second) and interactions for ten emotion categories and liking when evaluated by 26 participants. Numbers in bold indicate p-values < 0.1

		Context (DF=1)		Beer (DF=1)		Amount (DF=1)		Sample position (DF=1)		Context * Beer (DF=1)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.00	0.961	0.22	0.638	0.05	0.822	4.38	**0.038**	2.64	0.106
	Bored	3.63	**0.058**	3.85	**0.051**	4.46	**0.036**	0.28	0.598	0.31	0.576
	Content	6.67	**0.011**	0.65	0.423	0.05	0.816	4.88	**0.028**	0.98	0.323
	Excited	5.51	**0.020**	1.04	0.310	0.14	0.704	0.14	0.708	1.24	0.267
	Nostalgic	1.00	0.318	0.43	0.512	0.14	0.712	0.46	0.496	2.87	**0.092**
	Disconfirmed	1.01	0.316	1.37	0.244	1.08	0.299	9.41	**0.002**	0.96	0.329
	Disgusted	0.01	0.907	2.48	0.117	0.39	0.532	4.16	**0.043**	0.98	0.323
	Tame/safe	0.83	0.364	0.00	0.992	1.39	0.239	2.89	**0.091**	0.11	0.736
	Underwhelmed	1.63	0.203	0.81	0.369	0.80	0.373	0.45	0.504	0.17	0.677
	Curious	3.82	**0.052**	0.08	0.779	1.30	0.256	1.89	0.171	0.37	0.546
	Liking	2.80	**0.096**	0.52	0.471	0.00	0.956	5.81	**0.017**	0.65	0.421

Table continues on the next page

Table 5.4 continued

		Context * Amount (DF=1)		Context * Sample position (DF=1)		Beer *Amount (DF=1)		Beer * Sample position (DF=1)		Amount * Sample position (DF=1)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.03	0.872	0.19	0.663	0.53	0.468	5.01	**0.026**	0.77	0.382
	Bored	0.61	0.436	0.29	0.588	0.19	0.660	14.82	< **0.001**	0.73	0.395
	Content	0.72	0.398	0.80	0.373	0.04	0.835	3.32	**0.070**	0.02	0.880
	Excited	0.54	0.465	0.29	0.588	0.00	0.969	13.76	< **0.001**	0.08	0.781
	Nostalgic	0.00	0.985	0.11	0.736	0.02	0.885	1.27	0.261	0.34	0.558
	Disconfirmed	0.49	0.483	0.33	0.564	0.54	0.465	25.26	< **0.001**	0.58	0.447
	Disgusted	0.46	0.496	1.93	0.166	0.01	0.917	13.81	< **0.001**	0.10	0.756
	Tame/safe	0.15	0.699	1.60	0.208	0.41	0.522	0.22	0.640	2.74	0.099
	Underwhelmed	0.88	0.349	0.04	0.838	0.31	0.578	8.42	< **0.001**	0.00	0.982
	Curious	0.02	0.884	1.85	0.176	0.03	0.872	1.84	0.177	0.15	0.700
	Liking	0.66	0.417	0.20	0.659	0.26	0.609	22.99	< **0.001**	0.13	0.718

Table continues on the next page

Table 5.4 continued

		Context * Beer * Amount (DF=1)		Context * Beer * Sample position (DF=1)		Context * Amount * Sample position (DF=1)		Beer * Amount * Sample position (DF=1)		Context* Beer * Amount * Sample position (DF=1)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.02	0.881	4.52	**0.035**	0.44	0.507	0.09	0.768	0.03	0.855
	Bored	0.40	0.528	1.29	0.258	0.17	0.683	0.50	0.479	0.03	0.862
	Content	0.03	0.867	5.40	**0.021**	0.66	0.419	0.37	0.542	0.95	0.331
	Excited	0.01	0.943	7.26	**0.008**	0.05	0.826	0.03	0.863	2.56	0.111
	Nostalgic	0.33	0.564	4.25	**0.041**	0.04	0.845	0.02	0.888	0.27	0.601
	Disconfirmed	0.07	0.796	14.13	**0.000**	0.39	0.534	0.01	0.928	2.39	0.124
	Disgusted	0.09	0.767	11.32	**0.001**	0.01	0.911	0.00	0.961	1.36	0.245
	Tame/safe	0.07	0.798	6.54	**0.011**	0.20	0.655	0.11	0.739	0.06	0.814
	Underwhelmed	0.08	0.782	1.64	0.202	1.37	0.243	0.87	0.353	0.90	0.344
	Curious	0.34	0.559	0.01	0.937	0.00	0.950	0.03	0.860	2.41	0.122
	Liking	0.05	0.827	11.04	**0.001**	0.02	0.894	0.11	0.740	1.54	0.216

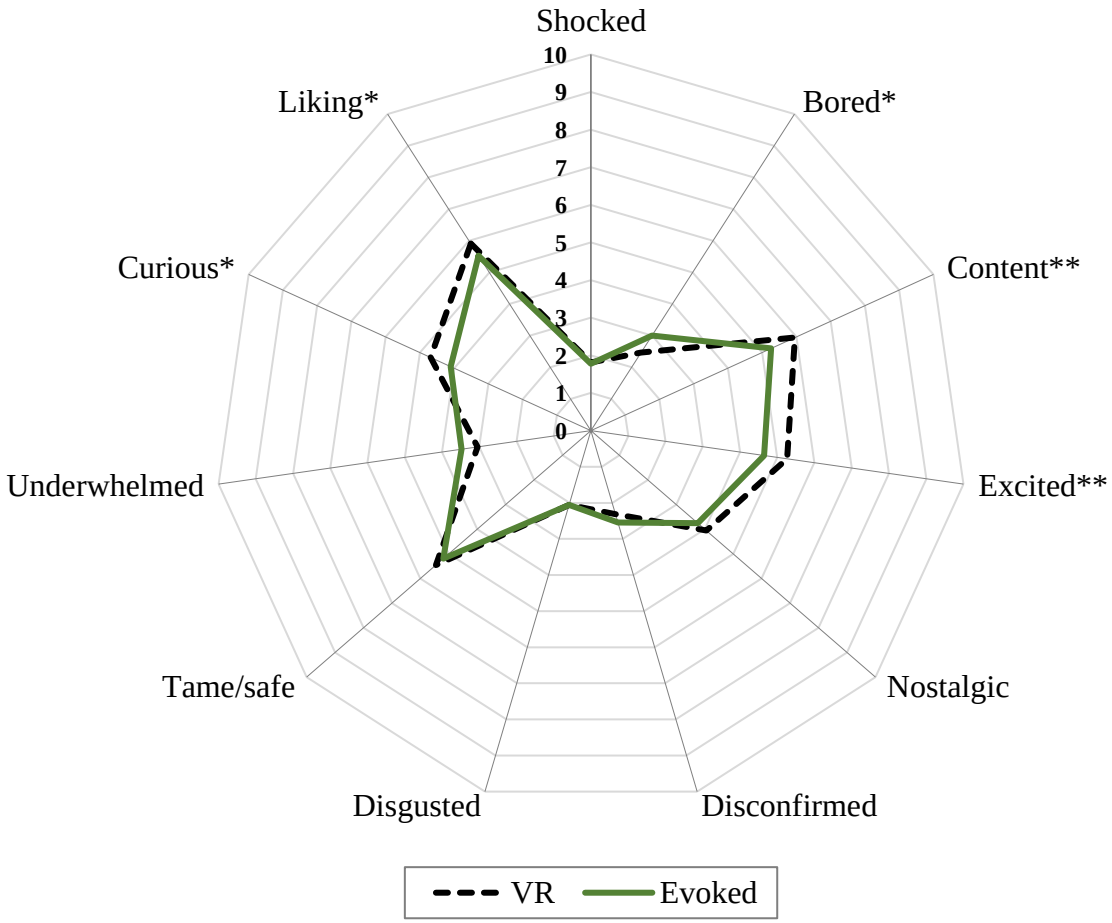


Figure 5.4: *Mean ratings for ten emotion categories and liking collected during the VR session and Evoked session, averaged over participants (N = 26) and beers (Ale and Lager).*

***Table 5.5**: Mean scores and Standard errors (SE) per context (VR Training session, VR session and Evoked session) for ten emotion categories and liking collected from 26 participants (over two beers after one sip and 200 ml)*

		VR Training		VR		Evoked	
		Mean	SE	Mean	SE	Mean	SE
Emotional response	Shocked	1.85 A	0.20	1.81 A	0.19	1.78 A	0.18
	Bored	2.26 B	0.22	2.47 AB	0.22	3.00 A	0.23
	Content	7.01 A	0.19	5.96 B	0.21	5.26 C	0.19
	Excited	6.38 A	0.21	5.26 B	0.22	4.65 C	0.19
	Nostalgic	4.60 A	0.25	4.06 AB	0.24	3.75 B	0.20
	Disconfirmed	2.04 A	0.23	2.31 A	0.22	2.55 A	0.23
	Disgusted	1.76 A	0.22	2.06 A	0.22	2.05 A	0.21
	Tame/safe	6.12 A	0.20	5.46 B	0.22	5.20 B	0.20
	Underwhelmed	2.66 B	0.25	3.05 AB	0.26	3.48 A	0.27
	Curious	5.95 A	0.24	4.69 B	0.24	4.10 B	0.20
	Liking	6.48 A	0.20	5.91 B	0.20	5.52 B	0.20

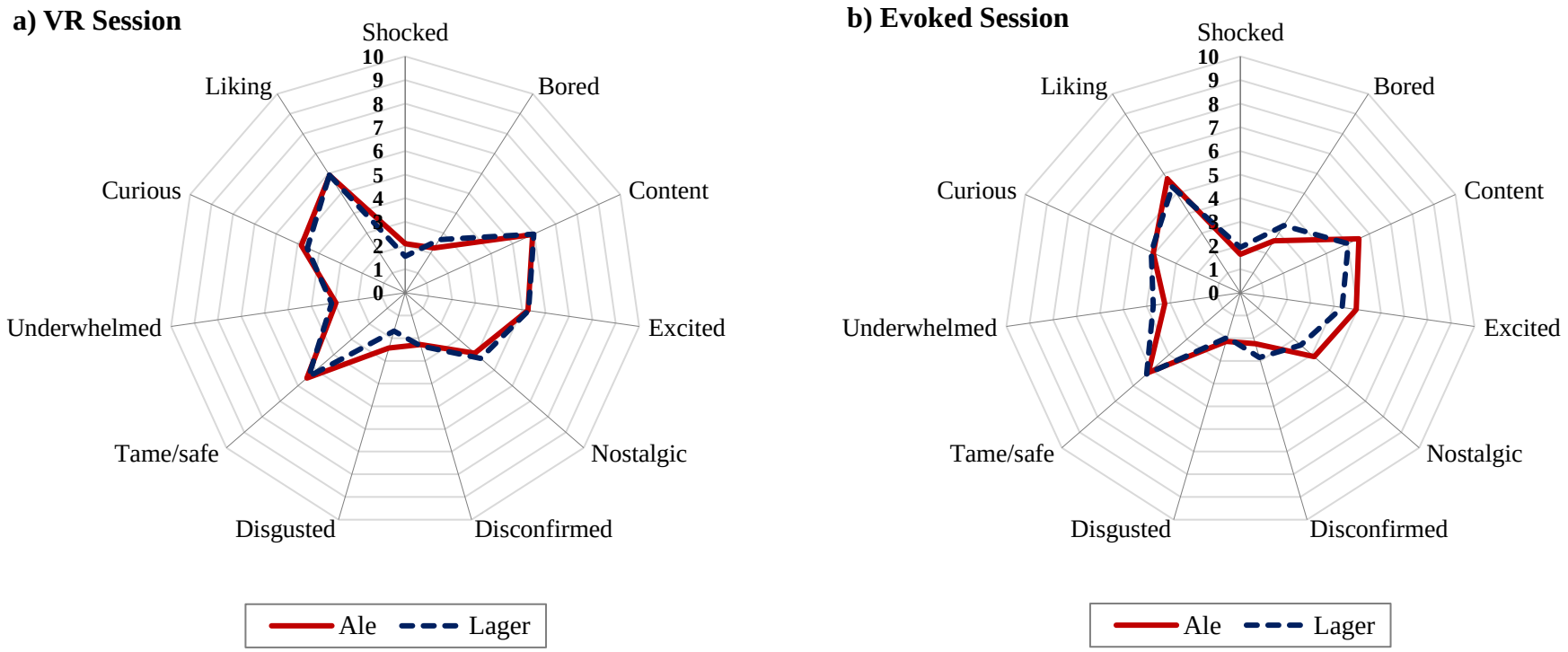


***Figure 5.5**: Mean ratings (N = 26) for ten emotion categories and liking per beer collected in a) VR session and the b) Evoked session respectively.*

5.3.1.2 The effect of VR Training

Table 5.6 shows the results for when the data from the Training session was included in the ANOVA, comparing the responses between the VR Training, VR session and Evoked session. As can be gathered from the low *p*-values (< 0.1) for session effect on seven emotion categories as well as liking in Table 5.6, responses differed across the VR Training session, VR session and Evoked session to a greater extent than between the VR and Evoked session (Table 5.4). Comparing the mean scores and post hoc groupings per session in Table 5.5, revealed that during the VR Training session, participants gave higher scores on Content, Excited, Tame/safe, Curious and Liking than during the VR session. This difference in response between the first (VR Training) and second time (VR session) participants were exposed to the beers in the VR context can also be seen when comparing the response profile from the VR Training session (Figure 5.6) to the response profile from the VR session (Figure 5.5a).

An even stronger variation in scores was found between the Training session and the Evoked session (Table 5.5). Response scores on liking and positive emotions Content, Excited, Nostalgic, Tame/safe and Curious were clearly higher during the VR Training than during the Evoked session. Negative emotion categories Bored and Underwhelmed were scored lower during the VR Training than during the Evoked session (Table 5.5).

The effects of beer, amount, sample position when the data from the VR Training session was included (Table 5.6) were similar to the findings from the analysis that only included the VR session and Evoked session (Table 5.4). However, the consumed amount appeared to have a stronger effect on Curious

3635 (p = 0.076, Table 5.6). Participants reported higher Curiosity after a single sip
3636 (Mean = 5.15, SE = 0.19) than after drinking approximately 200 ml (Mean =
3637 4.68, SE = 0.19). Furthermore, the interaction effect between context and beer
3638 on Nostalgic was no longer evident (p = 0.266, Table 5.6). Again, strong
3639 effects of Beer*Sample position and Context*Beer*Sample position were
3640 found while the other interactions did not reach p-values < 0.1 (Table 5.6).

***Table 5.6**: F and p-values from a four-factor ANOVA with session (VR Training, VR and Evoked), beer (Ale and Lager), consumed amount (one sip and 200 ml) and sample position (first or second) with interactions for ten emotion categories and liking when evaluated by 26 participants. Numbers in bold indicate p-values < 0.05*

a)		Session (DF=2)		Beer (DF=1)		Amount (DF=1)		Sample position (DF=1)		Session * Beer (DF=2)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.04	0.961	0.05	0.832	0.00	0.983	4.27	**0.040**	1.36	0.259
	Bored	3.51	**0.031**	5.78	**0.017**	4.76	**0.030**	0.16	0.688	0.16	0.856
	Content	21.17	< **0.01**	0.14	0.711	0.26	0.612	11.60	< **0.001**	0.89	0.410
	Excited	19.70	< **0.01**	0.74	0.389	0.25	0.617	1.67	0.198	0.72	0.486
	Nostalgic	3.41	**0.034**	0.93	0.337	0.01	0.929	1.18	0.278	1.33	0.266
	Disconfirmed	1.65	0.195	1.38	0.241	0.57	0.449	19.26	< **0.001**	0.49	0.615
	Disgusted	0.75	0.474	2.09	0.150	0.39	0.531	12.50	< **0.001**	0.68	0.506
	Tame/safe	5.27	**0.006**	0.00	0.970	1.88	0.172	4.66	**0.032**	0.06	0.943
	Underwhelmed	2.66	**0.072**	0.18	0.674	0.92	0.339	2.16	0.143	0.56	0.570
	Curious	17.52	< **0.01**	0.01	0.938	3.17	**0.076**	1.35	0.247	0.34	0.713
	Liking	6.90	< **0.01**	0.31	0.578	0.04	0.850	13.05	< **0.001**	0.40	0.674

Table 5.6 continued

b)		Session * Amount (DF=2)		Session * Sample position (DF=2)		Beer *Amount (DF=1)		Beer * Sample position (DF=1)		Amount * Sample position (DF=1)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.07	0.930	0.26	0.770	1.05	0.306	0.12	0.726	0.23	0.629
	Bored	0.47	0.629	1.24	0.290	0.13	0.714	15.35	< **0.001**	0.00	0.993
	Content	0.41	0.663	0.86	0.426	0.02	0.893	4.35	**0.038**	0.11	0.745
	Excited	0.26	0.768	0.85	0.428	0.04	0.850	13.76	< **0.001**	0.00	0.982
	Nostalgic	0.12	0.891	0.14	0.870	0.24	0.622	2.38	0.124	0.18	0.671
	Disconfirmed	0.45	0.639	0.75	0.473	1.93	0.166	16.22	< **0.001**	0.30	0.586
	Disgusted	0.24	0.785	2.10	0.124	0.01	0.911	8.49	< **0.001**	0.07	0.785
	Tame/safe	0.09	0.916	0.82	0.443	0.58	0.447	0.48	0.490	1.91	0.168
	Underwhelmed	0.45	0.636	0.44	0.646	0.58	0.448	5.67	**0.018**	0.11	0.736
	Curious	0.20	0.822	1.07	0.343	0.19	0.664	0.36	0.548	0.01	0.929
	Liking	0.33	0.719	0.66	0.516	0.60	0.441	16.70	< **0.001**	0.02	0.897

Table 5.6 continued

c)		Session * Beer * Amount (DF=2)		Session * Beer * Sample Position (DF=2)		Session* Amount * Sample position (DF=2)		Beer * Amount * Sample position (DF=1)		Session * Beer * Amount * Sample position (DF=2)	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Emotional response	Shocked	0.04	0.963	7.45	< **0.001**	0.52	0.597	0.03	0.855	0.04	0.957
	Bored	0.23	0.798	1.23	0.293	1.15	0.319	0.82	0.366	0.02	0.984
	Content	0.17	0.844	2.79	**0.063**	0.35	0.702	0.55	0.460	0.49	0.615
	Excited	0.06	0.943	4.03	< **0.001**	0.15	0.861	0.00	0.964	1.26	0.286
	Nostalgic	0.26	0.774	1.97	0.142	0.08	0.920	0.01	0.929	0.13	0.879
	Disconfirmed	0.31	0.734	10.17	< **0.001**	0.31	0.737	0.01	0.928	1.11	0.332
	Disgusted	0.10	0.905	7.97	< **0.001**	0.02	0.983	0.00	0.946	0.66	0.517
	Tame/safe	0.03	0.966	4.95	**0.008**	0.54	0.581	0.04	0.847	0.08	0.927
	Underwhelmed	0.04	0.957	2.26	0.107	0.83	0.437	0.38	0.537	0.74	0.480
	Curious	0.22	0.803	1.01	0.367	0.30	0.742	0.02	0.895	1.13	0.325
	Liking	0.05	0.951	7.73	< **0.001**	0.10	0.907	0.24	0.628	0.73	0.482

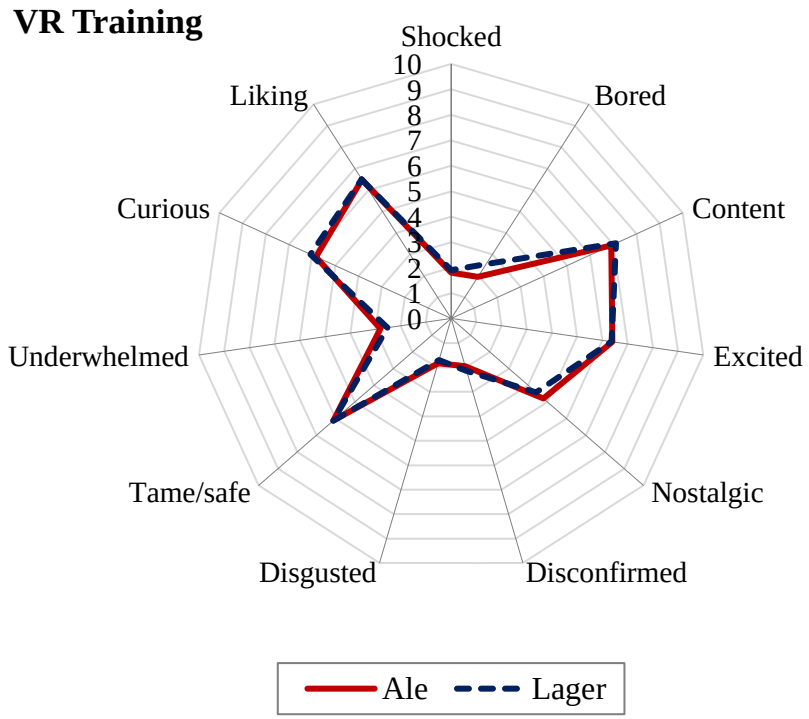


***Figure 5.6**: Mean ratings for ten emotion categories and liking per beer collected in the VR Training session.*

5.3.2 Participant engagement

5.3.2.1 Engagement questionnaire

Engagement was measured among all 27 participants at the end of the VR session and the Evoked session. Factor analysis on the 20 items related to engagement (Table 5.2) gave a two-factor solution that explained 56.2% of the variance. Most items loaded strongly on the first factor (loadings between 0.459 and 0.920) except item 5 and item 20 which loaded heavy on factor 2 (-0.421 and -0.738 respectively). Not a single item loaded negatively on the first factor, which can be seen as a justification for the reverse coding of some of the items. Cronbach's alpha was 0.94 for total engagement when including all twenty items. The dimensions Environmental Aesthetics, Involvement, Sensory Awareness and Realism showed high internal consistency (Cronbach's alpha > 0.8), while Immersion and Distraction and in particular Novelty scored low on internal consistency (Cronbach's alpha < 0.6, Table 5.7). The total engagement score was higher in the VR session than in the Evoked session ($p < 0.001$) and, as clearly shown in Table 5.7, the VR session also scored higher than the Evoked context on each individual dimension of engagement ($p < 0.001$).

***Table 5.7**: Mean scores and standard errors of all items from the engagement questionnaire rated by 27 participants during the Evoked and VR session. F and p-values on context effect were obtained through two-way mixed model ANOVA on dimension scores using participant as random factor and context condition as fixed factor.*

			Average				Context effect (DF=1)	
			Evoked		VR			
	Dimension	**Item**	***M***	***SE***	***M***	***SE***	***F***	***p***
strongly disagree (-3) to strongly agree (+3)	1. Usability	1. The testing environment assisted in my evaluations of the beers	0	0.25	1.48	0.2	20.08	< 0.001
	2. Environmental Aesthetics	Total Environmental Aesthetics (Cronbach's α = 0.84)	-0.56	0.27	1.8	0.15	62.72	< 0.001
		2. The testing environment was appealing	-0.85	0.3	1.74	0.17	58.6	< 0.001
		3. The testing environment engaged my senses	-0.26	0.32	1.85	0.17	35.29	< 0.001
	3. Novelty	Total Novelty (Cronbach's α = 0.04)	0.33	0.2	1.61	0.21	17.06	< 0.001
		4. The testing environment incited my curiosity	-0.59	0.31	2.11	0.23	50.49	< 0.001
		5. The testing environment distracted me [R]	1.26	0.28	1.11	0.3	0.13	0.72
	4. Involvement	Total Involvement (Cronbach's α = 0.81)	0.01	0.21	1.9	0.17	34.47	< 0.001
		6. The testing environment was boring [R]	-0.37	0.3	1.52	0.23	18.6	< 0.001
		7. The testing environment was fun	-0.74	0.28	1.93	0.27	40.26	< 0.001
		8. I was engaged in the sensory task I performed	1.15	0.19	2.26	0.16	14.29	< 0.001
	5. Immersion	Total Immersion (Cronbach's α = 0.55)	-1.04	0.22	1.33	0.17	80.83	< 0.001
		9. I felt like I was in a bar	-1.74	0.28	1.59	0.22	95.12	< 0.001
		10. I lost track of time	-0.33	0.32	1.07	0.24	17.27	< 0.001

[R] *Indicates items that were reverse-coded.*

Table 5.7 continued

			Average				Context effect (DF=1)	
			Evoked		VR			
not at all (1) to very much (7)	6. Sensory awareness	Total sensory awareness (Cronbach's α = 0.82)	3.49	0.19	5.56	0.13	83.38	< 0.001
		11. How completely were all of your senses engaged by the testing environment?	3.41	0.26	5.48	0.15	52.2	< 0.001
		12. How much did the visual aspects of the testing environment involve you?	2.33	0.21	5.3	0.21	108.19	< 0.001
		13. How much did the auditory aspects of the testing environment involve you?	4.74	0.24	5.89	0.19	15.58	< 0.001
	7. Realism	Total Realism (Cronbach's α = 0.83)	2.96	0.18	5.04	0.16	70.68	< 0.001
		14. How disconnected did you feel from the testing environment? [R]	3.48	0.28	4.52	0.29	5.14	0.03
		15. How much did your experiences in the testing environment seem consistent with real-world experiences?	2.67	0.32	5.07	0.26	48.14	< 0.001
		16. How completely did you feel immersed in the testing environment?	2.70	0.20	5.19	0.21	52.24	< 0.001
		17. How involved were you in the testing environment experience?	3.00	0.21	5.37	0.19	93.25	< 0.001
	8. Distraction	Total Distraction (Cronbach's α = 0.48)	4.22	0.18	5.6	0.15	31.18	< 0.001
		18. How aware were you of events occurring in the real world around you? [R]	2.89	0.32	5.19	0.31	29.94	< 0.001
		19. How quickly did you adjust to the testing environment experience?	4.74	0.26	5.96	0.19	10.42	< 0.001
		20 How much did the testing environment interfere or distract you from performing your sensory evaluation? [R]	5.04	0.25	5.67	0.23	4.47	< 0.001
	9. Total Engagement Score (Cronbach's α = 0.94)		**9.43**	**1.22**	**24.32**	**0.82**	**83.82**	**< 0.001**
	10. Comfort	21. Did you experience dizziness?	1.11	0.08	1.22	0.08	1.3	0.26
		22. Did you experience discomfort in this test environment?	1.37	0.21	1.59	0.18	1	0.33
	11. Drinking beer	23. Did you experience problems with consuming the beer?	1.56	0.23	2.37	0.26	9.31	0.01

[R] *Indicates items that were reverse-coded.*

5.3.2.2 Novelty

The VR condition was considered more novel than the Evoked condition ($p < 0.001$, Table 5.7). The majority of participants had no or very limited prior experience with VR and therefore participants were expected to have a reaction to the novelty of their experience with the VR bar. The first phrases a participant expressed after removal of the headset were captured during the interview at the end of the training session. Sometimes a participant would spontaneously express their reaction but in most cases the interviewer probed for a response using questions in Table 5.3.

Figure 5.7 shows a word cloud of the words spoken by participants after taking off the headset. The most frequently expressed term was ‘really’ which was used in combination with adjectives such as ‘good’, ‘weird’, ‘cool’ and ‘strange’ which were also expressed frequently. The use of the words ‘weird’ and ‘strange’ seems to convey novelty while ‘good’ and ‘cool’ seem to express enjoyment. The use of these terms confirms the high level of novelty and associated enjoyment that participants seemed to experience during the VR Training session.



Figure 5.7: *Word cloud of first few phrases participants spoke after taking off the headset, excluding stop words.*

5.3.2.3 Environment and realism

The VR testing environment was considered more appealing in terms of Environmental Aesthetics than the Evoked condition during which participants were sat in sensory booths ($p < 0.001$, Table 5.7). During the interview several participants said that the VR Bar felt like a normal pub and that it made them feel more relaxed and comfortable. The environment was easily recognisable as the Student Union Bar and familiar to the participants. Most participants confirmed that it was a typical environment that they could find themselves in in real life, although some participants mentioned that they would have chosen to sit somewhere else in the bar. When asked to describe the environment most participants spoke about the people that were present in the VR Bar*: "Quite realistic. Different groups of people, so to the left you'd see two guys, you had a couple on a date and then you had a group of friends, which was quite realistic. And they were, because I've been to [the Student Union Bar] before, kind of where they would sit. If that makes sense. The date probably would be on there and then the groups tend to be at the back more. So that was quite realistic."*

Overall, perceived realism was higher for the VR context than the evoked condition ($p < 0.001$, Table 5.7). Participants varied in their responses as to which elements they thought made the VR Bar more or less realistic. The quality of the visuals and technical difficulties due to practical limitations of VR (which will be discussed further in section 5.3.3) were the main features that made the experience less realistic, whilst audio, the presence of people and the time spent in VR appeared to increase realism. Familiarity with the bar seemed to have both a positive and negative effect on perceived realism.

The image quality was mentioned by nearly half of all participants. The 360° footage was described to be blurry and less clear than the modelled table and other objects that were closer to the participant. The modelled objects were describe as cartoony and game-like. The design of the bottles specifically was described as simplistic and strange because it appeared as if the bottles were still closed and there was no liquid inside the bottles. As one participant described: *"Obviously you know that you are in virtual reality. Like you know that some things are going to not be as real as they are in real life so you don't really pay attention to that. At some point I was like oh they are closed but then you just drink from it and it is fine".* Another limitation that was noticed by some participants was the double-layered floor where the sheer computer-modelled floor covered the floor of the 360° video. Some participants noted that this was unrealistic and distracting. One of the participant that noticed the floor said: *"The quality of the picture wasn't like fantastic. It wasn't like crystal clear. Yeah so maybe a little bit pixilated around me. The floor was kind of slightly hallucinogenic. It was kind of moving and a bit changing colour. You could see everything clearly, but it just wasn't kind of like… I know probably you're never going to get that, but it wasn't like you were fully immersed in it because your eyes... it just wasn't the same quality."* Interestingly, few participants looked down and therefore many did not notice the layered floor. Other technical limitations included 'glitches' and problems with object tracking, where objects suddenly made impossible movements or appeared to be in a different place than their physical counterparts. This made drinking beer more challenging than participants would experience in real life. Some participants expressed that they found it odd that they could not see

themselves. The fact that they could not see their arms was related to difficulties with drinking the beers and thus reduced realism. One participant in particular expressed that they were disappointed that they could not see themselves in VR because they had done VR gaming with a full haptic body suit before and therefore knew it was possible in VR. Interestingly however, quite a few participants did not mention 'being invisible' as an aspect that made the situation less realistic, and when asked about it they expressed that they had not noticed it or did not find it strange because they normally would not look at themselves if they were in a bar anyway. Only one participant seemed impressed that the VR Bar was 360° whereas none of the other participants articulated amazement about the fact that the experience was projected all around them. This seems related to this single participant's expectations based on previous experience with playing a PlayStation VR game which was not 360°.

The presence of people in the bar appeared to be important for perceived realism. Several participants expressed that the behaviour and placement of the people in the VR bar were as they would expect in real life. Having noises and conversations of people in the background was also found to make the bar setting more realistic, although the music was at a lower level than some participants would expect in the bar. Some participants mentioned that it seemed quiet or unnaturally calm because none of the people were moving around in the bar. Audio seemed to increase the level of perceived realism as well as immersion. Results from the engagement questionnaire revealed that the auditory aspects in VR involved participants more than in the Evoked condition ($p < 0.001$, Table 5.7). Not only was hearing other people and noises

from the bar environment said to make the experience more realistic, participants also mentioned how the sound drowned out any noises from the physical environment. It is interesting to note that in the evoked condition the visual aspects of the testing environment were less involving than in VR ($p < 0.001$, Table 5.7). Some participants specifically expressed that they were usually very aware of their physical surroundings when doing a test in the sensory booths.

The longer participants were exposed to the VR environment, the more realism seemed to increase: *"I think once you're over the initial feeling of like... that you know what you are doing, I think it did become quite realistic. Like you are just in a quiet bar and there is a few people dotted around. And like I even started tuning in to like what was being said. Which you would naturally do in a bar anyway if you start hearing people's conversations. So I found that as I became more comfortable with what I was actually doing and like I say I'm getting to the point where I was daydreaming and stuff, it felt quite like a normal environment."* More time spent in VR also facilitated participants to get used to things that initially made the experience less realistic, such as the appearance of objects and visuals and not being able to see their own arms. As one participant stated with regards to being invisible: *"To be honest, after about 30 seconds I forgot about it."* Some participants said they were still conscious of being part of an experiment and that there was a researcher in the room, whereas other participants claimed that they momentarily forgot where they physically were, to the point where they forgot the lay-out of the physical room they were in. One participant commented on the fact that he did not feel strange about drinking beer in the afternoon because he was immersed in VR:

"I'd certainly say in terms of the fact that I thought I was drinking a beer and didn't feel awkward drinking a beer at 2:30[pm] felt quite nice. In like a random lab, the fact that it took me out of that was really cool."

The fact that all participants were familiar with the actual bar that the 360° video was based on seemed to increase the sense of realism for some, whilst for others it highlighted aspects in which the VR bar differed from the real bar. One participant for whom realism increased said: *"I guess also it being somewhere I recognised makes it more realistic. Rather than being in a situation you have never been in before. Then it is kind of forced."* On the other hand some participants seemed more critical of certain aspects of the VR environment that differed from their real life experience in the bar. Some participants mentioned that the bar is usually much busier and the music much louder when they visit it. One participant commented on the lack of menus, which would normally be present on the tables. Two participants mentioned that they did not see any staff present in the bar.

5.3.2.4 Distraction

The engagement questionnaire revealed higher scores for the Distraction dimension in VR than under the evoked condition ($p < 0.001$, Table 5.7) which indicated that participants felt less distracted in VR than when they were exposed to sounds and pictures in the sensory booths. During the interview participants said they focussed on the table when doing the questionnaire in VR and did not feel distracted by the environment. The focus was more on the environment around them when participants were drinking the beer in VR: *"...when I wasn't drinking I was focussing on the table and focussing on the lines [line scales] and things like that. I wasn't really focussing on the*

environment when I was doing the questionnaire. But when I was drinking the beer I was focussing more on the environment then. So I don't think I found it really distracting at all." Some participants commented that because there was nothing distracting going on it felt less like being in a real bar environment. In a few instances participants did not immediately notice instructions that appeared on the table because they were looking around. The novelty of the VR experience was also mentioned as a distracting factor. Technical limitations of the VR environment could also be distracting: *"The only distracting thing I would say is that... just because sometimes things were a little off in terms of where the bottle was or compared to where it was in reality. Not by a huge amount but enough to be like yeah this isn't quite right."*

5.3.3. Practical implications of VR

5.3.3.1 Questionnaire

The method of rating emotion categories and liking using the pen and the line scale depicted on the virtual table was described by participants as clear, easy and intuitive. Specifically physically moving their arm in VR to give ratings was positively received and participants articulated feeling more engaged with the task. This seemed to be in line with the higher score on Usability and Involvement for VR compared to the Evoked condition in the engagement questionnaire ($p < 0.001$, Table 5.7). Several participants said that they felt less removed from the environment while doing the questionnaire and preferred the VR method over using a computer or phone: *"You feel a little bit removed when you are doing it on a phone or a computer. Like it is a bit more formal. Whereas this is far more informal."* Some participants also reacted positively to seeing a score appear on the scale: *"It actually felt quite satisfying to do. I*

think particularly that, you know, I was just kind of gesturing in the air, but it was actually putting it where... putting the cross in the place where I was trying to put it. So yeah there was something quite satisfying about that. Perhaps a little more difficult to put things exactly where you were although I think the fact that there was a number actually helped to know exactly what I was kind of saying. Whereas that swayed when there wasn't a number [during sessions in the sensory booths]. It was kind of more of a guess. You know, it was easier to kind of correct for what you meant compared to what you put down almost." Despite the overall positive response to the questionnaire, there were also some negative comments. One participant expressed a need to focus to get comfortable with the questionnaire method. Another participant thought the pen was a bit heavy because of the tracking device attached to it. One participant remarked that she focussed on the lower end of the scale because she felt that that part of the scale was located more in front of her. Some participants commented on the ambiguity of the emotion words themselves or difficulties with expressing how they felt because they forgot immediately after drinking the beer. Sometimes the technology itself was not working properly, which caused some frustration with participants: *"[The questionnaire was] easy. Very easy and very friendly. I only had a little bit of a moaning moment, a bit of a concern, when I was tapping the table and the cross was not coming on it. But I think technology itself sometimes can have its problems."* Some participants started out by tapping the physical table with the pen, but stopped doing that once they realised their score was registered without physical contact between the pen and the table. Several participants explicitly expressed that they preferred to simply hover above the table to make a score:

"I did not really want the contact, because I accidentally hit the table once and I was like whoops. And that kind of took me out of it. I think just being able to lightly tap is probably enough"

5.3.3.2 Drinking beer

At the end of the engagement questionnaire participants were asked to report if they had any difficulties consuming the beer. As can be seen in Table 5.7, low levels of problems with consuming the beer were reported in both conditions, but were higher in VR than in the Evoked condition ($p = 0.01$, Table 5.7). As mentioned in section 5.3.2.3, the closed appearance of the virtual bottles reduced the experienced realism. Interestingly, none of the participants said they were reluctant to drink the beer from the virtual bottles. Some participants mentioned their disliking of the beers and that it was unnatural or challenging to drink a beer for the sake of the experiment that they normally would not have, indicating reduced ecological validity due to the experimental situation rather than due to VR exposure. Drinking the beer from the bottles while wearing the VR headset received both comments related to ease and difficulty with drinking. Several participants described difficulties with drinking the beer from the bottles. Examples included struggles with picking up the bottle, getting the top of the bottle to their mouth and problems with knocking the bottle against the headset or having to lift up their head more to be able to tilt the bottle enough to drink. These difficulties were caused by technical limitations of the VR technology such as the fact that participants could not see their own arms, nor the beer in the bottle, as well as misalignment of the physical bottle and the VR-modelled bottle. Participants described being more careful and slow in their movements at first and getting more confident with

handling the objects with practice. Participants also described not handling the bottles as they normally would, for example feeling where the neck and opening of the bottle were with one hand to aid the coordination of the bottle to their mouths. Several participants described holding the bottle in their hands while they would normally put it down on the table. The reason they gave for holding the bottle was fear of 'losing it' and having problems with picking it up again. Participants also claimed that they drank the beers faster than they normally would because of a lack of social interaction, not usually drinking beer from the bottle, boredom because they could not take out their phone as they normally would do in a bar, and because they were still conscious of being part of an experiment, which they were eager to complete in a timely manner. The fact that some participants did not put the beer down, but held it in their hands was also mentioned as a reason for drinking faster, because it stimulated them to take sips more frequently. One participant mentioned being worried about missing the 'stop drinking sign' because tilting his head to take sips meant losing sight of the table. As a consequence the participant said to take smaller sips than usual to make sure he did not drink too much. In fact, several participants articulated worries about drinking more than they were allowed because they could not see the amount of beer that was left in the bottle. However, most participants reported that they got more comfortable with handling the bottles over time.

5.3.3.3 Social context

The relevance of social context for the scenario of having a drink in a bar was strongly confirmed by participants who said they would always meet up with someone if they went to a bar. Most said it would be unusual for them to sit in

a bar by themselves and so the written scenario was helpful in that it provided an explanation to why they were sitting alone. All participants confirmed that the written scenario described a realistic situation that would be typically experienced by them in real life. Several participants said the scenario made them feel more at ease about the fact that they were sitting by themselves in a bar situation: *"That [scenario] certainly helped. Because, you know, it would be a bit weird I suppose if you were just saying right, you're in a bar, you're on your own and you are basically doing nothing except drinking beer. I wouldn't do that, but I would expect to be meeting someone so that was quite useful, because it helped to sort of set the scene."* Although the scenario was helpful it also had limitations. The questionnaire distracted away from the scenario because the task is not something they would do if they were waiting for a friend. Several participants expressed that they would probably take out their phone instead. Some participants mentioned that the friend would have been very late if they had to have two beers waiting from them. Participants explained that they engaged in people-watching and eavesdropping onto other people's conversations just as they would do in real life. In other words the presence of other people made the experience more interesting and seemed to stimulate natural behaviour and increase the perception of a normal bar experience. One participant mentioned that she noticed one of the people in the VR Bar looking at her at some point and that made her feel more like she was really there, because she was seen by people around here like you normally would be in a bar. Another however described turning around to have a look at what was going on behind her, whereas in real life she would not have done that so obviously to look at other people. Some participants mentioned feeling

‘lonely’, or that they missed being able to engage in social interaction with people in the VR bar, or having a friend present to talk to.

5.3.3.4 Comfort

As part of the engagement questionnaire participants were asked to report on their level of comfort. As can be seen in Table 5.7, similarly low average scores were found for both context conditions ($p > 0.1$). In line with the results from the engagement questionnaire participants reported minimal levels of discomfort during the interviews. The familiarity of the setting seemed to make participants feel more comfortable and relaxed, especially after a few minutes of exposure to the VR environment*: “I think it does feel like you are having a drink in a place where you would more normally have a drink in that sort of respect. And I did feel very relaxed. I think the more I got used to everything the more I relaxed. So I probably just relaxed into it the longer I was wearing the headset.”*

Sometimes their experience was more comfortable than expected, like this participant who expressed she felt more at ease in VR than in the sensory booths: *“I would have thought that I would be more uncomfortable than I was. But I felt really comfortable in it, and it sounds really stupid but it was easier to like daydream. When you're in a sensory booth, I find that you notice your surroundings and you don't normally daydream but with that I was sort of thinking, like I normally would”*. Even though they did not feel uncomfortable, several participants did say that they ‘noticed’ the feeling of the headset pressing on their face and that this reminded them that the experience was not real. Two participants described the headset as ‘big’ and ‘chunky’ and one mentioned that it was quite warm to wear the headset. Some participants

experienced problems because the headset wasn't optimally adjusted to their head. The headset could be adjusted for the participant's head size and when this was not done adequately, participants experienced discomfort because the headset would slide on their face. In some cases participants described that incorrect setting of the headset lead to blurry (peripheral) vision in VR. The hygiene cover was also a cause for minor discomfort during some training sessions. In two cases the hygiene cover got stuck to the participant's glasses when the headset was removed but this was experienced as comical rather than uncomfortable. During one session the hygiene cover slipped down a bit when participant accidentally hit the headset with a beer bottle and she described how she 'just tucked it up' and was fine again after that. Only one participant mentioned experiencing dizziness but not until after the headset was taken off because of the readjustment to reality. The earphones were commented on to be on-ear rather than over-ear which made them slightly less comfortable and the overall experience less immersive.

5.4. Discussion

5.4.1 Emotional response and liking

The results showed an overall effect of VR on emotional response and liking to beer. When participants were exposed to the VR condition they reported lower scores for Bored and Underwhelmed and higher scores on positive emotion categories Content, Excited, Nostalgic and Curious, as well as liking. No previous studies have measured emotional response on continuous line scales in VR compared to measurement in a lab setting and so this finding cannot be compared to previous research. However, higher hedonic ratings in an immersive room compared to a controlled setting were found by Bangcuyo et

al. (2015) and Hathaway and Simons (2017) and (Sinesio et al., 2018). Sinesio et al. (2019) on the other hand did not find an effect of context on liking when comparing an immersive room, a 360° video, a VR modelled environment, a standard sensory lab setting and a real bar environment. Furthermore, Sinesio et al. (2019) found better product discrimination between beer products on a check-all that apply (CATA) list of emotions in a CLT setting than in a real pub, VR 360° video, VR model and immersive room. In their study the beers were not differentiated on liking in any of the contexts (Sinesio et al., 2019). Further research is needed to determine the effect of the VR context used in the current study on product differentiation.

Evidence was found that exposure to a single sip of beer made participants feel more Curious and less Bored than consuming a larger amount of up to 200 ml. This finding discourages the use of single sip testing and in favour of presenting participants with more representative amount of products in sensory consumer testing, in line with earlier findings (Chapter 3 and Chapter 4).

During the first exposure to VR (the VR Training session) participants showed higher scores on liking and emotion categories associated with positive affect and lower on some of the emotion categories related to negative affect than when they were exposed to the VR condition a second time (during the VR session). This finding indicates that the novelty of the VR experience had a strong effect on the response to beer. During a second exposure these effects were significantly reduced.

5.4.2 Engagement

The engagement questionnaire was found to be a useful tool for the purpose of measuring participant engagement, although the internal consistency of some of the dimensions was low. The current study found low Cronbach's alpha for Novelty, Immersion and Distraction. Certain items in these dimensions possibly need to be reconsidered. A standardised method to measure engagement would be beneficial for comparisons of participant engagement across context studies.

Qualitative data from participant interviews showed that VR clearly had a positive impact on total participant engagement, which was in line with the levels of engagement that were quantified with the engagement questionnaire. This finding was in line with previous research that found that VR enhances participant engagement compared to a method using pictures and participants' imagination (Andersen et al., 2019), and in broader sense immersive technologies have been shown to increase engagement compared to standard lab settings (Bangcuyo et al., 2015; Hathaway & Simons, 2017; Sinesio et al., 2019).

5.4.3 Novelty

Interestingly, the second exposure to the VR context (the VR session) yielded more similar emotional responses to those measured during the Evoked session, which could mean that the more positive emotional response was due to the high level of novelty related to exposure to VR technology. This indicates that the VR training should not only be considered as a way to train participants on the test protocol, but also as a strategy to reduce the impact of novelty on resulting emotional responses. The frequent use of the words

‘weird’, ‘strange’, ‘good’ and ‘cool’ participants used to express how to they felt about their first exposure to the VR Bar seemed to illustrate both novelty and enjoyment. However, it should also be considered that the participants saw the training session as practice and therefore might not have been as committed to correctly evaluating their true emotional responses as much as trying to get a grasp on the situation and correctly handling the objects. In addition, these effects could partially be due to repeated exposure to the products as a result of the within-subject test design. Therefore more research is needed to determine if this is a true effect.

It is interesting to consider whether novelty has an effect on product differentiation and if the reduction of novelty leads to stronger product differentiation on emotional response. Unfortunately the current study did not include sufficient participants to allow valid analyses to determine product differentiation within each context condition. Sinesio et al. (2018) found that simulating context in an immersive room led to higher hedonic scores and reduced product differentiation compared to a sensory laboratory setting. More research is needed to study the effects of novelty related to new technologies on hedonic ratings and emotional response.

The results show signs of how prior experience with VR impacted participants’ expectations. The fact that all participants except one had no specific reaction to the fact that the experience was projected all around them could be an indication that they either expected the experience to be 360°, or that they had no expectations at all due to a lack of prior experience with VR. Future studies should consider that as immersive technology becomes more common, more and more participants will be familiar with VR, which will alter their

expectations and thus their perceived realism. A clear example of this was found in this study, where one participant that had experienced VR gaming with a haptic body suit was very disappointed with the fact that she could not see her arms in the VR bar. Other participants that had no such prior VR experience either did not notice or quickly accepted the fact that they were invisible, which could indicate that participants are more forgiving of elements that reduce realism if they have no prior experiences to compare it with.

5.4.4 Immersion and realism

When evoking context the aim is to simulate the natural consumption experience as closely as possible. As with any simulation, imagination plays a role with VR as well as with the Evoked condition used in this study. The more sensory inputs presented in VR, the easier it is for participants to visualize and feel immersed in the virtual environment (Cowan & Ketron, 2019), and the more closely responses approach those obtained in real life situations (Stelick & Dando, 2018). Stimulation of all five senses (i.e., vision, taste, touch, smell and hearing) can make virtual experiences more immersive (Cowan & Ketron, 2019). Olfaction was not included in the current experiment and potentially could have enhanced participants' sense of immersion further. Although both VR and the Evoked condition involved visual and auditory stimuli, results from the engagement questionnaire revealed that participants were less involved with the visual and auditory aspects during the evoked condition than in VR. Liu et al. (2019) found evidence that in an immersive room visual cues were attended to the most, followed by auditory cues. In the current study it seems that, based on the ratings for item twelve and thirteen in the engagement questionnaire, within the Evoked condition auditory aspects were more

involving than visual aspects. Possibly this was because participants actively had to direct their attention to the tablet device in their sensory booth to see the slideshow of pictures from the Student Union bar. Andersen et al. (2019) found that an imagined context evoked with pictures was less engaging than VR. The exposure to audio recordings used in the current study might have provided added advantage in terms of the engagement with the context. However, since immersion was clearly lower in the Evoked condition participants likely had to rely more heavily on their imagination than in the VR condition. The quality of the visuals and technical difficulties in VR were the main features that made the experience less realistic, while audio seemed to increase realism.

Immersion and realism seemed to be related. Although there were no direct questions in the interview script on immersion, the line of questioning naturally led participants to express statements related to their level of immersion, mostly when talking about realism or when they described their experience with drinking the beers. Some participants described feeling self-conscious and drinking faster because they were still conscious of being part of an experiment and that there was a researcher in the room, whereas other participants claimed that they momentarily forgot where they physically were, to the point where they forgot the lay-out of the physical room they were in. Participants expressed being conscious of being part of an experiment and so anything that reminded them of the fact that they were not really in the bar reduced their perception of realism and sense of immersion. The image quality of the 360° video was frequently mentioned as an aspect that reduced realism. Similar to findings by Sinesio et al. (2019) this study found problems with image distortion on the peripheral of the 360° footage. Technological

limitations might be overcome as technology advances. There are however certain things that can be done to optimise realism. The use of a 360° camera that captures high-quality footage in addition to better lighting in the bar when the footage was captured could have improved the image quality in the current study and should be considered for future studies.

The presence of people in the VR Bar was found to be a key element that enhanced realism. The captured images of real people behaving naturally in a way that is expected in a bar strongly impacted the atmosphere of the environment. Since the same footage was used in both the VR Training and VR session, any eye-catching behaviour such as the walking around of participants was avoided. Although this was perceived as unnatural by some participants it was also believed to minimise distraction. Certain details such as the presence of bar staff and menus on the tables could have enhanced realism further.

The time spent in VR also impacted realism and immersion. The longer participants were exposed to the VR Bar, the more their sense of realism and immersion increased. For this reason it is advisable to allow adequate time for participants to experience the VR environment and adjust to it before starting any evaluation task. Further research is needed to determine the ideal exposure time and if this is the same over different and familiar versus unfamiliar contexts. The use of a dummy sample to increase exposure time before evaluating the product of interest might offer a good solution whilst at the same time reducing first-order effects which are known to impact emotion data (Dorado, Pérez-Hugalde, et al., 2016).

Familiarity with the bar seemed to have both a positive and negative effect on perceived realism. Participants' personal history impacts how they interpret and experience a contextual setting (Andersen et al., 2019; Köster, 2003). In the current study we tried to account for this effect by only recruiting participants that had been in the specific bar that was featured the 360° video. Based on participants' responses during their interviews, all participants recognized the bar as a familiar environment and for a proportion of the participants this was found to increase the level of realism and immersion and helped to make them feel more comfortable. Delarue et al. (2019) speculated that consumers used their own context-related memories to project themselves in their 'own world'. Similarly we believe that for some participants certain incorrect details of the environment such as the absence of bar staff and the lack of loud music were overwritten by participants' memory filling in missing details. This is in line with Paris et al. (2011) who found that when a context where cigarette cues are normally present was simulated in VR, this environment elicited significant cravings even when specific cigarette cues were absent. For other participants the familiarity with the bar, perhaps combined with the fact that they were aware of being in an experimental setting made them more critical or aware of details that differed between the VR bar and their real life experience.

5.4.5 Distraction

Although minimising distraction enhances engagement with the evaluation task (Bangcuyo et al., 2015), a certain level of external noise was found to be important for the realism of the bar environment. Immersive technologies have been shown to induce similarly higher levels of distraction compared to

sensory laboratory setting as a real pub environment (Sinesio et al., 2019). In the current study however, the VR condition was found to be less distracting than the Evoked condition, indicating that participants felt more focussed on the task of evaluating their responses to the beers while they were immersed in the VR bar than when they were exposed to auditory and visual cues on the tablet in the sensory booths. Similarly, Bangcuyo et al. (2015) found participants were less distracted in a virtual coffee house in an immersive room than under standard sensory lab conditions.

5.4.6 Consuming beer in VR

The current study made use of an innovative method in which object tracking allowed the participant to stay fully immersed during consumption. After training all participants were able to consume the beers without any assistance from the researcher. Being able to consume products without taking off the headset is crucial for the context to have effect, since participants quickly adjust to reality after exposure to VR (Andersen et al., 2019). An additional advantage is that sample presentation can be truly blind and the appearance of the samples can be manipulated in any way (Stelick & Dando, 2018). Disadvantages included the current limitations of the technology such as inaccuracies, glitches and virtual objects making impossible movements, which reduced realism. Tracking devices add weight and bulk to objects and manipulation of the objects requires practice. If participants were to cover the trackers with their hands the tracking accuracy would reduce further. Additionally, the fact that participants could not see their own hands and arms was a disadvantage (Stelick & Dando, 2018). As a result these technical limitations could impact participants' natural consumption behaviour, which in

turn might impact sensory perception and emotional response. For example, holding beer in hands could lead to a more rapid increase in temperature than putting the beer down between sips. Sensory perception of beer can change at a function of temperature due to increased volatility of flavour compounds. Beer temperature has also been shown to impact emotional response (Dorado, Chaya, et al., 2016; Yang et al., 2018). Due to the disadvantages of object tracking, other technologies need to be considered and explored, such as hand-tracking (Ung, Menozzi, Hartmann, & Siegrist, 2018), augmented reality glasses or mixed reality systems (Flavián, Ibáñez-Sánchez, & Orús, 2019), where a virtual environment (either modelled environment, a 360° video or both) is combined with actual real-time display of the physical environment through a camera that is mounted on the headset (Andersen et al., 2019; Korsgaard, Nilsson, & Bjørner, 2017). However, studies using such systems also found problems with food and beer consumption (Korsgaard, Bjøner, & Nilsson, 2019; Sinesio et al., 2019) and thus more work is required to find the best approach.

5.4.7 Questionnaire design

Considering the brief nature of emotions (Coppin & Sander, 2016) and the fact that participants appeared to quickly re-adjust from VR to physical reality after removal of the headset (Andersen et al., 2019) emotion measurement during context exposure is strongly advisable. Questionnaire design for sensory consumer research in VR is still under-developed (Stelick & Dando, 2018). The current study explored an innovative approach whereby a questionnaire was integrated in design objects in the virtual environment. Using a tracked marker pen to scale emotion categories in VR was found to be easy and

intuitive for participants. Furthermore, the fact that participants felt more involved with the questionnaire can be considered an advantage of VR over traditional questionnaire methods and opens up considerable possibilities in terms of improving engagement with questionnaires. Other researchers should be encouraged to embrace the use of more creative data collection methods that make the task more stimulating for research participants. Since rating all ten emotion categories and liking after exposure to the beer still requires memory, it would be interesting to explore simultaneous self-report measurement and product consumption.

5.4.8 Written scenario

In the current study the use of a written scenario prior to context exposure to reduce the impact of a lack of social interaction in VR was explored, as previous research has shown that written scenarios can help participants to imagine different contexts strong enough to impact emotional response (Dorado, Chaya, et al., 2016; Piqueras-Fiszman & Jaeger, 2014b). The written scenario was positively received by participants, although some participants described feeling lonely due to the lack of social interaction. Something similar was found in previous research where participants in a VR beach context reported feelings of being 'secluded' (Andersen et al., 2019). Thus the lack of social interaction can make participants feel isolated, which is a clear disadvantage of VR compared to immersive rooms, when simulating contexts that are usually associated with social interaction. There is clearly still a need for more advanced developments to allow realistic social interaction in VR (Stelick & Dando, 2018). Based on the results of the current study it is recommended to use a written scenario to provide participants with an

explanation of how or why they find themselves in the situation simulated with immersive technology. In other words, to provide context to the context.

5.4.9 Comfort

In the current study participants seemed to experience minimal levels of discomfort. The only discomfort that was reported was related to the headset, which could mostly be prevented by adequately adjusting the headset to the participants head and thus it is recommended to take the time to adjust headsets to participants when simulating context in VR for consumer research purposes.

5.5. Conclusion

This study explored the effect of simulating a consumption context with VR on participant engagement when measuring hedonic and emotional responses to beer. In addition, practical implications of VR related to measuring responses, consuming beverages, social context and comfort were studied.

Evidence was found that high levels of novelty related to a first exposure to VR led to a predominantly positive emotional response and high liking compared to a second exposure. A training session prior to a data collection session in VR was observed to reduce novelty and associated effects, resulting in responses that were more similar to those measured when evoking context with pictures and sound recordings in sensory booths.

Compared to a less immersive approach using pictures and sound recordings (Evoked condition), VR showed clear advantages in terms of participant engagement. Key elements were identified that improved realism and immersion such as audio, the time spent in VR and the realistic simulation of

having other people present in the environment. Image quality of the 360° footage and technical limitations were found to reduce participants' sense of realism and immersion, while familiarity with the environment that was simulated showed both enhanced and reduced realism and immersion.

This study explored an innovative method of using tracking devices to facilitate beer consumption and the integration of a questionnaire in virtual design objects. Using tracking devices to facilitate beer consumption from glass beer bottles allowed participants to consume products independently, while maintaining full immersion in VR. However, evidence was found that the use of tracking devices affected normal drinking behaviour. Therefore future research should look to technological advances in minimising the impact of such tracking devices. The integrated questionnaire in VR was found to benefit usability and enhance participant engagement with the task of reporting responses using line scales. The possibilities of using VR to facilitate more interactive questionnaire methodologies seem very promising for increasing engagement. Providing participants with a written scenario prior to VR exposure was found to be useful as a way to account for the lack of social interaction in VR. However, approaches to improve social interaction should also be considered. VR did not seem to cause participants discomfort in this study.

Despite the technological limitations of VR that were observed, the findings affirm that VR can be used as a tool to simulate context in sensory consumer research. This paper contributes to improved methodologies in sensory consumer research by highlighting relevant considerations when using VR in sensory consumer tests.

Chapter 6

6. Conclusions

The core objective of this research was to determine the effects of context on the measurement of consumers' emotional response to beer for the purpose of sensory consumer research and to explore new ways of improving contextual factors in a controlled setting. Emotion and context are complex phenomena that consist of multiple components. In the introduction emotion was defined as "*an event-focussed, two-step, fast process consisting of (1) relevance-based emotion elicitation mechanisms that (2) shape a multiple emotional response"* (Coppin & Sander, 2016), and the research focussed on measuring the experiential 'feeling' component of the multiple emotional response with self-report questionnaires. Context was defined as *"a set of events and experiences that are not part of the reference event that one is interested in but have some relationship to it"* (Rozin & Tuorila, 1993), and whilst it was acknowledged that there are seven categories of contextual factors ('Who', 'What', 'How', 'Where', 'When', 'With whom' and the 'Task') that are always present in a sensory consumer research situation, the research mainly focussed on comparing real-life and controlled test environments, and exploring methods to bring more realistic consumption contexts into controlled settings.

The research confirmed variability in emotion between contexts. In Chapter 2, emotional response collected in a natural consumption environment for beer (i.e. a bar) showed stronger product differentiation than the same responses, collected from the same participants in a sensory laboratory. It was demonstrated that individuals can differ in the degree to which their emotional responses vary across situations, and that evoking context in a sensory

laboratory resulted in an emotional response that was more similar to the real bar context with consumers that were context-sensitive. These findings indicate that emotional responses to beer obtained in a sensory laboratory provide low validity for emotional responses in a bar environment, but can be improved by evoking a relevant consumption context, which is valuable information for sensory consumer research test design. Furthermore, clustering participants on liking in order to better understand their emotional responses can be recommended based on this research.

In Chapter 3, responses given after a single sip mildly differed from responses collected after consumption of a representative quantity of beer (i.e. a glass) for some emotion categories that were measured in this research. Generally participants reported higher curiosity after a single sip and feeling more content after the consumption of a representative quantity. In addition, responses collected after consumption of glass of beer were found to have a higher test-retest reliability than responses collected after a single sip. Test-retest reliability was comparable between a controlled setting and natural consumption environment, thus indicating that measuring emotional responses to beer in a real-life consumption environment, where there is little experimental control, does not necessarily reduce stability of the test results. Furthermore, the emotional response to a lager beer was found to be stable across repeated exposures over the course of several weeks. A training session, used as a dummy session, reduced first-order session effects on emotional response during the studies in Chapter 2 and Chapter 4, demonstrating that the task of self-reporting emotions can itself impact emotional responses. Presenting participants with representative amounts of beer and testing the

same beer more than once for sensory consumer research can be recommended based on these findings.

Collecting responses through online questionnaires via participants' smartphones is a convenient way to collect emotion data in natural consumption environments. In Chapter 4 scoring line scales with a finger on the small-sized screen of a phone was found to be less accurate and therefore likely to underestimate differences in emotional response compared to using a mouse and larger display monitor of a desktop computer. Sensory consumer researchers should therefore be aware of the risk that existing differences might not be detected when smartphones are used for data collection.

In Chapter 5 Virtual Reality (VR) technology was found to be a promising tool to simulate natural consumption context and to study context effects in a controlled sensory research environment. VR was used to realistically immerse participants in a simulated consumption context and resulted in a higher participant engagement compared to evoking context with picture and sound recordings in a sensory booth. However, a first exposure to VR was associated with high levels of novelty which resulted in relatively high scores on positive emotion categories in response to beer. During a second exposure to the VR environment the positive emotion categories were scores less high in response to the beer, and more similar to the levels measured in a sensory laboratory setting. The research uncovered key elements that improved realism and immersion in VR such as audio, the time spent in VR, providing participants with a scenario prior to VR exposure, and the realistic simulation of having other people present in the environment. Image quality and technical limitations were found to reduce participants' sense of realism and immersion.

The use of tracking devices was successful as a means of facilitating independent sample consumption by participants in VR, however, their use did appear to affect normal drinking behaviour. Using the same tracking technology to allow participants to score emotion categories with a marker pen on the digital table was found to be easy and intuitive for participants to use and demonstrated the added value to participant engagement that creative questionnaire solutions in VR can offer. This research uncovered relevant considerations when using VR in sensory consumer tests, which contributes to improved methodologies in sensory consumer research.

Further questions still remain regarding the measurement of consumers' emotional responses in context. Below, suggestions are made for topic of future research:

- In this research, only one natural consumption environment for beer was investigated, namely a student union bar. However, in real life consumers drink beer in a variety of different situations and so it is of interest to investigate how much emotional responses to the same beers vary across different consumption situations. Future research could compare responses between home use tests, concerts, festivals, restaurants etc. as well as exploring the effects across different bars.
- Any natural consumption context comprises many different factors such as music, lighting, decorations, aroma, social context etc. In the bar that was used as natural consumption environment for this research these factors were not controlled. However, it is not known how these specific elements impact emotional responses to beer. Further research could be conducted to untangle effects of specific factors that make up

a natural consumption context, in order to determine which elements should be included in a consumer test design that aims to optimise ecological validity. This type of information will allow sensory scientists to make informed decisions to optimise context in their research projects.

- Comparisons were made across context conditions based on the product differentiation of an Ale and Lager style beer, and therefore additional research is needed to confirm the findings on product differentiation with beers in the same style. Sensory consumer researchers often aim to differentiate between similarly liked products. For example, future research could compare emotional responses across several lagers in different consumption situations.
- In Chapter 3 the effects of repeated emotional response measurement were studied for a single beer. However, repeated exposure might have an effect on product differentiation between different beers in a sensory consumer test. Boredom with the products or with the emotion measurement questionnaire itself could lead to a loss of participants' motivation and lead them to report less accurate responses. In other words, participants might just click through the questionnaire without considering their answers. Less accurate emotion measurement is likely to result in finding fewer differences between products in terms of the emotional responses they evoke. A study designed to repeatedly measure emotional responses to several products would be beneficial for uncovering the effect of repeated exposure on product differentiation with self-reported emotional response.

- The research in Chapter 5 showed that simulating a context with Virtual Reality technology had a strong impact on the measured emotional responses to beer. One of the limitations of this work was the number of participants, which was too low to draw any conclusions on the extent to which the emotional responses differentiated the two beers. The excitement related to exposure to novel technology might reduce the accuracy of self-reported emotional responses. Therefore, quantitative research with a representative number of consumers (N = 100) is required to determine the effects of simulating context with VR on product differentiation with emotional response.

References

Achar, C., So, J., Agrawal, N., & Duhachek, A. (2016). What we feel and why we buy: the influence of emotions on consumer decision-making. *Current Opinion in Psychology, 10*, 166-170. doi:http://dx.doi.org/10.1016/j.copsyc.2016.01.009

Alaoui-Ismaïli, O., Robin, O., Rada, H., Dittmar, A., & Vernet-Maury, E. (1997). Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation. *Physiol Behav, 624*. doi:10.1016/s0031-9384(97)90016-0

Allison, R. I., & Uhl, K. P. (1964). Influence of Beer Brand Identification on Taste Perception. *Journal of Marketing Research, Vol. 1* (3), pp. 36-39. doi:10.2307/3150054

Andersen, I. N. S. K., Kraus, A. A., Ritz, C., & Bredie, W. L. P. (2019). Desires for beverages and liking of skin care product odors in imaginative and immersive virtual reality beach contexts. *Food Res Int, 117*, 10-18. doi:10.1016/j.foodres.2018.01.027

Aquilani, B., Laureti, T., Poponi, S., & Secondi, L. (2015). Beer choice and consumption determinants when craft beers are tasted: An exploratory study of consumer preferences. *Food Quality and Preference, 41*, 214-224. doi:http://dx.doi.org/10.1016/j.foodqual.2014.12.005

Asioli, D., Varela, P., Hersleth, M., Almli, V. L., Olsen, N. V., & Næs, T. (2017). A discussion of recent methodologies for combining sensory and extrinsic product properties in consumer studies. *Food Quality and Preference, 56*, 266-273. doi:http://dx.doi.org/10.1016/j.foodqual.2016.03.015

Bangcuyo, R. G., Smith, K. J., Zumach, J. L., Pierce, A. M., Guttman, G. A., & Simons, C. T. (2015). The use of immersive technologies to improve consumer testing: The role of ecological validity, context and engagement in evaluating coffee. *Food Quality and Preference, 41*, 84-95. doi:10.1016/j.foodqual.2014.11.017

Barthomeuf, L., Rousset, S., & Droit-Volet, S. (2009). Emotion and food. Do the emotions expressed on other people's faces affect the desire to eat liked and disliked food products? *Appetite, 52*(1), 27-33. doi:http://dx.doi.org/10.1016/j.appet.2008.07.002

Bell, R., Meiselman, H. L., Pierson, B. J., & Reeve, W. G. (1994). Effects of adding an Italian theme to a restaurant on the perceived ethnicity, acceptability, and selection of foods. *Appetite, 22*(1), 11-24. doi:10.1006/appe.1994.1002

Bell, R., & Pliner, P. L. (2003). Time to eat: the relationship between the number of people eating and meal duration in three lunch settings. *Appetite, 41*(2), 215-218. doi:http://dx.doi.org/10.1016/S0195-6663(03)00109-0

Berridge, K., & Winkielman, P. (2003). What is an unconscious emotion?(The case for unconscious "liking"). *Cognition and Emotion, 17*(2), 181-211. doi:10.1080/02699930302289

Beyts, C., Chaya, C., Dehrmann, F., James, S., Smart, K., & Hort, J. (2017). A comparison of self-reported emotional and implicit responses to aromas in beer. *Food Quality and Preference, 59*, 68-80. doi:10.1016/j.foodqual.2017.02.006

Bhumiratana, N., Adhikari, K., & Chambers, E. (2014). The development of an emotion lexicon for the coffee drinking experience. *Food Research International, 61*, 83-92. doi:10.1016/j.foodres.2014.03.008

Birch, L. L., Billman, J., & Richards, S. S. (1984). Time of day influences food acceptability. *Appetite, 5*(2), 109-116. doi:http://dx.doi.org/10.1016/S0195-6663(84)80030-6

Boutrolle, I., & Delarue, J. (2010). The effects of context on liking: implications for hedonic measurements in new product development-8. In S. R. Jaeger (Ed.),

Consumer-Driven Innovation in Food and Personal Care Products. (pp. pp.175-218).
Boutrolle, I., Delarue, J., Arranz, D., Rogeaux, M., & Koster, E. P. (2007). Central location test vs. home use test: Contrasting results depending on product type. *Food Quality and Preference, 18*(3), 490-499. doi:10.1016/j.foodqual.2006.06.003
Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle Reflex Modification: Emotion or Attention? *Psychophysiology, 27*(5), 513-522. doi:10.1111/j.1469-8986.1990.tb01966.x
Canetti, L., Bachar, E., & Berry, E. M. (2002). Food and emotion. *Behavioural Processes, 60*(2), 157-164. doi:http://dx.doi.org/10.1016/S0376-6357(02)00082-7
Cardello, A. V., Meiselman, H. L., Schutz, H. G., Craig, C., Given, Z., Lesher, L. L., & Eicher, S. (2012). Measuring emotional responses to foods and food names using questionnaires. *Food Quality and Preference, 24*(2), 243-250. doi:10.1016/j.foodqual.2011.12.002
Carvalho, F. R., Moors, P., Wagemans, J., & Spence, C. (2017). The Influence of Color on the Consumer's Experience of Beer. *Frontiers in Psychology, 8*(2205). doi:10.3389/fpsyg.2017.02205
Chaya, C., Eaton, C., Hewson, L., Vazquez, R. F., Fernandez-Ruiz, V., Smart, K. A., & Hort, J. (2015). Developing a reduced consumer-led lexicon to measure emotional response to beer. *Food Quality and Preference, 45*, 100-112. doi:10.1016/j.foodqual.2015.06.003
Chrea, C., Grandjean, D., Delplanque, S., Cayeux, I., Le Calvé, B., Aymard, L., . . . Scherer, K. (2009). Mapping the semantic space for the subjective experience of emotional responses to odours. *Chemical Senses [Peer Reviewed Journal], 34*(1), 49-62 doi:10.1093/chemse/bjn052
Churchill, A., & Behan, J. (2010). Comparison of methods used to study consumer emotions associated with fragrance. *Food Quality and Preference, 21*(8), 1108-1113. doi:http://dx.doi.org/10.1016/j.foodqual.2010.07.006
Coppin, G., & Sander, D. (2016). 1 - Theoretical Approaches to Emotion and Its Measurement - Meiselman, Herbert L. In *Emotion Measurement* (pp. 3-30): Woodhead Publishing.
Cowan, K., & Ketron, S. (2019). A dual model of product involvement for effective virtual reality: The roles of imagination, co-creation, telepresence, and interactivity. *Journal of Business Research, 100*, 483-492. doi:https://doi.org/10.1016/j.jbusres.2018.10.063
Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General, 145*(7), 830-843. doi:10.1037/xge0000172
Cruwys, T., Bevelander, K. E., & Hermans, R. C. J. (2015). Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite, 86*, 3-18. doi:http://dx.doi.org/10.1016/j.appet.2014.08.035
Dalenberg, J. R., Gutjar, S., ter Horst, G. J., de Graaf, K., Renken, R. J., & Jager, G. (2014). Evoked Emotions Predict Food Choice. *PLoS ONE, 9*(12), e115388. doi:10.1371/journal.pone.0115388
Danner, L., Ristic, R., Johnson, T. E., Meiselman, H. L., Hoek, A. C., Jeffery, D. W., & Bastian, S. E. P. (2016). Context and wine quality effects on consumers' mood, emotions, liking and willingness to pay for Australian Shiraz wines. *Food Res Int, 89*(Pt 1), 254-265. doi:10.1016/j.foodres.2016.08.006
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. doi:https://doi.org/10.1016/j.foodqual.2013.01.004

De Graaf, C., Cardello, A. V., Matthew Kramer, F., Lesher, L. L., Meiselman, H. L., & Schutz, H. G. (2005). A comparison between liking ratings obtained under laboratory and field conditions: the role of choice. *Appetite, 44*(1), 15-22. doi:10.1016/j.appet.2003.06.002

De Graaf, C., Kramer, F. M., Meiselman, H. L., Lesher, L. L., Baker-Fulco, C., Hirsch, E. S., & Warber, J. (2005). Food acceptability in field studies with US army men and women: relationship with food intake and food choice after repeated exposures. *Appetite, 44*(1), 23-31. doi:10.1016/j.appet.2004.08.008

De Wijk, R. A., He, W., Mensink, M. G. J., Verhoeven, R. H. G., & De Graaf, C. (2014). ANS Responses and Facial Expressions Differentiate between the Taste of Commercial Breakfast Drinks. *PLoS ONE, 9*(4), e93823. doi:10.1371/journal.pone.0093823

De Wijk, R. A., Kooijman, V., Verhoeven, R. H. G., Holthuysen, N. T. E., & De Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. *Food Quality and Preference, 26*(2), 196-203. doi:https://doi.org/10.1016/j.foodqual.2012.04.015

Delarue, J., Brasset, A. C., Jarrot, F., & Abiven, F. (2019). Taking control of product testing context thanks to a multi-sensory immersive room. A case study on alcohol-free beer. *Food Quality and Preference, 75*, 78-86. doi:https://doi.org/10.1016/j.foodqual.2019.02.012

Deliza, R., & MacFie, H. J. H. (1996). The Generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: a review. *Journal of Sensory Studies, 11*(2), 103-128. doi:10.1111/j.1745-459X.1996.tb00036.x

Delplanque, S., Grandjean, D., Chrea, C., Coppin, G., Aymard, L., Cayeux, I., . . . Scherer, K. (2009). Sequential Unfolding of Novelty and Pleasantness Appraisals of Odors: Evidence From Facial Electromyography and Autonomic Reactions. *Emotion, 9*(3), 316–328. doi:10.1037/a0015369

Den Uijl, L. C., Jager, G., De Graaf, C., Waddell, J., & Kremer, S. (2014). It is not just a meal, it is an emotional experience – A segmentation of older persons based on the emotions that they associate with mealtimes. *Appetite, 83*, 287-296. doi:http://dx.doi.org/10.1016/j.appet.2014.09.002

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. *Advances in Consumer Research, 27*(1), 111-117.

Desmet, P. M. A., & Schifferstein, H. N. (2008). Sources of positive and negative emotions in food experience. *Appetite, 50*(2-3), 290-301. doi:10.1016/j.appet.2007.08.003

Dijksterhuis, G. B. (2016). New product failure: Five potential sources discussed. *Trends in Food Science & Technology, 50*, 243-248. doi:https://doi.org/10.1016/j.tifs.2016.01.016

Dorado, R., Chaya, C., Tarrega, A., & Hort, J. (2016). The impact of using a written scenario when measuring emotional response to beer. *Food Quality and Preference, 50*, 38-47. doi:10.1016/j.foodqual.2016.01.004

Dorado, R., Pérez-Hugalde, C., Picard, A., & Chaya, C. (2016). Influence of first position effect on emotional response. *Food Quality and Preference, 49*, 189-196. doi:https://doi.org/10.1016/j.foodqual.2015.12.009

Dzardanova, E., Kasapakis, V., & Gavalas, D. (2018). On the Effect of Social Context in Virtual Reality: An Examination of the Determinants of Human Behavior in Shared Immersive Virtual Environments. *IEEE Consumer Electronics Magazine, 7*(4), 44-52. doi:10.1109/MCE.2018.2816204

Eaton, C., Chaya, C., Smart, K. A., & Hort, J. (2018). Comparing a full and reduced version of a consumer-led lexicon to measure emotional response to beer. *Journal of Sensory Studies, 0*(0), e12481. doi:10.1111/joss.12481

Eaton, C., Chaya, C., Smart, K. A., & Hort, J. (2019). Comparing a full and reduced version of a consumer-led lexicon to measure emotional response to beer. *Journal of Sensory Studies, 34*(2), e12481. doi:10.1111/joss.12481

Edwards, J. S. A., Hartwell, H. J., & Brown, L. (2013). The relationship between emotions, food consumption and meal acceptability when eating out of the home. *Food Quality and Preference, 30*(1), 22-32. doi:http://dx.doi.org/10.1016/j.foodqual.2013.04.004

Edwards, J. S. A., Hartwell, H. J., & Giboreau, A. (2016). Emotions Studied in Context: The Role of the Eating Environment. In H. L. Meiselman (Ed.), *Emotion Measurement* (pp. 377-404). Oxford: Woodhead Publishing 2016.

Edwards, J. S. A., Meiselman, H. L., Edwards, A., & Lesher, L. (2003). The influence of eating location on the acceptability of identically prepared foods. *Food Quality and Preference, 14*(8), 647-652. doi:10.1016/S0950-3293(02)00189-1

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129. doi:10.1037/h0030377

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. *Journal of Personality & Social Psychology, 53*(4), 712-717.

Feldman Barrett, L. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science, 1*(1), 28-58. doi:10.1111/j.1745-6916.2006.00003.x

Feldman Barrett, L. (2016). Navigating the Science of Emotion. In H. L. Meiselman (Ed.), *Emotion Measurement* (pp. 31-63): Woodhead Publishing.

Feldman Barrett, L. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(11), 1833-1833. doi:10.1093/scan/nsx060

Feldman Barrett, L., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The Experience of Emotion. *Annual Review of Psychology, 58*(1), 373-403. doi:doi:10.1146/annurev.psych.58.110405.085709

Ferrarini, R., Carbognin, C., Casarotti, E. M., Nicolis, E., Nencini, A., & Meneghini, A. M. (2010). The emotional response to wine consumption. *Food Quality and Preference, 21*(7), 720-725. doi:http://dx.doi.org/10.1016/j.foodqual.2010.06.004

Fisher, C. M., King, S. K., Castura, J. C., & Findlay, C. J. (2016). Does Data Collection Device Affect Sensory Descriptive Analysis Results? *Journal of Sensory Studies, 31*(4), 275-282. doi:10.1111/joss.12210

Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2019). The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of Business Research, 100*, 547-560. doi:https://doi.org/10.1016/j.jbusres.2018.10.050

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations Among Emotion, Appraisal, and Emotional Action Readiness. *Journal of Personality and Social Psychology, 57*(2), 212-228. doi:10.1037/0022-3514.57.2.212

Galiñanes Plaza, A., Delarue, J., & Saulais, L. (2019). The pursuit of ecological validity through contextual methodologies. *Food Quality and Preference, 73*, 226-247. doi:https://doi.org/10.1016/j.foodqual.2018.11.004

Gendron, M., Roberson, D., Van der Vyver, J. M., & Feldman Barrett, L. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion, 14*(2), 251-262. doi:10.1037/a0036052

Giacalone, D., Frøst, M. B., Bredie, W. L. P., Pineau, B., Hunter, D. C., Paisley, A. G., . . . Jaeger, S. R. (2015). Situational appropriateness of beer is influenced by product familiarity. *Food Quality and Preference, 39*, 16-27. doi:http://dx.doi.org/10.1016/j.foodqual.2014.06.012

Gibson, E. L. (2006). Emotional influences on food choice: Sensory, physiological and psychological pathways. *Physiology & Behavior, 89*(1), 53-61. doi:http://dx.doi.org/10.1016/j.physbeh.2006.01.024

Gomez-Corona, C., Chollet, S., Escalona-Buendia, H. B., & Valentin, D. (2017). Measuring the drinking experience of beer in real context situations. The impact of affects, senses, and cognition. *Food Quality and Preference, 60*, 113-122. doi:10.1016/j.foodqual.2017.04.002

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. doi:10.1037/0022-3514.74.6.1464

Gutjar, S., Dalenberg, J. R., De Graaf, C., De Wijk, R. A., Palascha, A., Renken, R. J., & Jager, G. (2015). What reported food-evoked emotions may add: A model to predict consumer food choice. *Food Quality and Preference, 45*, 140-148. doi:10.1016/j.foodqual.2015.06.008

Gutjar, S., De Graaf, C., Kooijman, V., De Wijk, R. A., Nys, A., Ter Horst, G. J., & Jager, G. (2015). The role of emotions in food choice and liking. *Food Research International, 76*, 216-223. doi:10.1016/j.foodres.2014.12.022

Hathaway, D., & Simons, C. T. (2017). The impact of multiple immersion levels on data quality and panelist engagement for the evaluation of cookies under a preparation-based scenario. *Food Quality and Preference, 57*, 114-125. doi:https://doi.org/10.1016/j.foodqual.2016.12.009

Hedderley, D. I., & Meiselman, H. L. (1995). Modelling meal acceptability in a free choice environment. *Food Quality and Preference, 6*(1), 15-26. doi:http://dx.doi.org/10.1016/0950-3293(94)P4208-N

Hein, K. A., Hamid, N., Jaeger, S. R., & Delahunty, C. M. (2010). Application of a written scenario to evoke a consumption context in a laboratory setting: Effects on hedonic ratings. *Food Quality and Preference, 21*(4), 410-416. doi:10.1016/j.foodqual.2009.10.003

Hein, K. A., Hamid, N., Jaeger, S. R., & Delahunty, C. M. (2012). Effects of evoked consumption contexts on hedonic ratings: A case study with two fruit beverages. *Food Quality and Preference, 26*(1), 35-44. doi:10.1016/j.foodqual.2012.02.014

Hersleth, M., Mevik, B. H., Naes, T., & Guinard, J. X. (2003). Effect of contextual factors on liking for wine - use of robust design methodology. *Food Quality and Preference, 14*(7), 615-622. doi:10.1016/S0950-3293(02)00190-8

Hersleth, M., Monteleone, E., Segtnan, A., & Naes, T. (2015). Effects of evoked meal contexts on consumers' responses to intrinsic and extrinsic product attributes in dry-cured ham. *Food Quality and Preference, 40*, 191-198. doi:10.1016/j.foodqual.2014.10.002

Hersleth, M., Ueland, O., Allain, H., & Naes, T. (2005). Consumer acceptance of cheese, influence of different testing conditions. *Food Quality and Preference, 16*(2), 103-110. doi:10.1016/j.foodqual.2004.02.009

Holthuysen, N. T. E., Vrijhof, M. N., De Wijk, R. A., & Kremer, S. (2017). "Welcome on board": Overall liking and just-about-right ratings of airplane meals in three different consumption contexts—laboratory, re-created airplane, and actual airplane *Journal of Sensory Studies, 32*(2), e12254. doi:doi:10.1111/joss.12254

Hort, J., Redureau, S., Hollowood, T., Marciani, L., Eldeghaidy, S., Head, K., . . . Taylor, A. J. (2008). The effect of body position on flavor release and perception: Implications for fMRI studies. *Chemosensory Perception, 1*(4), 253-257. doi:10.1007/s12078-008-9034-0

Izard, C. E. (2009). Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues. *Annual Review of Psychology, 60*(1), 1-25. doi:10.1146/annurev.psych.60.110707.163539

Jaeger, S. R., Cardello, A. V., Chheang, S. L., Beresford, M. K., Hedderley, D. I., & Pineau, B. (2017). Holistic and consumer-centric assessment of beer: A multi-measurement approach. *Food Research International, 99*, 287-297. doi:https://doi.org/10.1016/j.foodres.2017.05.004

Jaeger, S. R., Cardello, A. V., & Schutz, H. G. (2013). Emotion questionnaires: A consumer-centric perspective. *Food Quality and Preference, 30*(2), 229-241. doi:http://dx.doi.org/10.1016/j.foodqual.2013.05.015

Jaeger, S. R., Hort, J., Porcherot, C., Ares, G., Pecore, S., & MacFie, H. J. H. (2017). Future directions in sensory and consumer science: Four perspectives and audience voting. *Food Quality and Preference, 56*, 301-309. doi:10.1016/j.foodqual.2016.03.006

Jaeger, S. R., & Porcherot, C. (2017). Consumption context in consumer research: methodological perspectives. *Current Opinion in Food Science, 15*, 30-37. doi:https://doi.org/10.1016/j.cofs.2017.05.001

Jager, G., Schlich, P., Tijssen, I., Yao, J., Visalli, M., De Graaf, C., & Stieger, M. (2014). Temporal dominance of emotions: Measuring dynamics of food-related emotions during consumption. *Food Quality and Preference, 37*, 87-99. doi:http://dx.doi.org/10.1016/j.foodqual.2014.04.010

Jiang, W., Niimi, J., Ristic, R., & Bastian, S. E. P. (2016). The Effects of Immersive Context and Wine Flavor on Consumer Wine Flavor Perception and Emotions Elicited. *American Journal of Enology and Viticulture, 68*(1). doi:10.5344/ajev.2016.16056

Jiang, Y., King, J. M., & Prinyawiwatkul, W. (2014). A review of measurement and relationships between food, eating behavior and emotion. *Trends in Food Science & Technology, 36*(1), 15-28. doi:http://dx.doi.org/10.1016/j.tifs.2013.12.005

Kaneko, D., Toet, A., Brouwer, A.-M., Kallen, V., & Van Erp, J. B. F. (2018). Methods for Evaluating Emotions Evoked by Food Experiences: A Literature Review. *Frontiers in Psychology, 9*, 911-911. doi:10.3389/fpsyg.2018.00911

Keesman, M., Aarts, H., Ostafin, B. D., Verwei, S., Hafner, M., & Papies, E. K. (2018). Alcohol representations are socially situated: An investigation of beverage representations by using a property generation task. *Appetite, 120*, 654-665. doi:10.1016/j.appet.2017.10.019

Kemp, S. E., Hollowood, T., & Hort, J. (2009). *Sensory Evaluation: A Practical Handbook*. Chichester: Wiley-Blackwell.

King, S. C., & Meiselman, H. L. (2010). Development of a method to measure consumer emotions associated with foods. *Food Quality and Preference, 21*(2), 168-177. doi:10.1016/j.foodqual.2009.02.005

King, S. C., Meiselman, H. L., & Carr, B. T. (2010). Measuring emotions associated with foods in consumer testing. *Food Quality and Preference, 21*(8), 1114-1116. doi:10.1016/j.foodqual.2010.08.004

King, S. C., Meiselman, H. L., & Carr, T. B. (2013). Measuring emotions associated with foods: Important elements of questionnaire and test design. *Food Quality and Preference, 28*(1), 8-16. doi:http://dx.doi.org/10.1016/j.foodqual.2012.08.007

King, S. C., Meiselman, H. L., Hottenstein, A. W., Work, T. M., & Cronk, V. (2007). The effects of contextual variables on food acceptability: A confirmatory study. *Food Quality and Preference, 18*(1), 58-65. doi:10.1016/j.foodqual.2005.07.014

King, S. C., Weber, A. J., Meiselman, H. L., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference, 15*(7-8), 645-653. doi:10.1016/j.foodqual.2004.04.010

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

Korsgaard, D., Bjøner, T., & Nilsson, N. C. (2019). Where would you like to eat? A formative evaluation of mixed-reality solitary meals in virtual environments for older adults with mobility impairments who live alone. *Food Research International, 117*, 30-39. doi:https://doi.org/10.1016/j.foodres.2018.02.051

Korsgaard, D., Nilsson, N. C., & Bjørner, T. (2017, 19-19 March 2017). *Immersive eating: evaluating the use of head-mounted displays for mixed reality meal sessions.* Paper presented at the 2017 IEEE 3rd Workshop on Everyday Virtual Reality (WEVR).

Köster, E. P. (2003). The psychology of food choice: some often encountered fallacies. *Food Quality and Preference, 14*(5-6), 359-373. doi:10.1016/S0950-3293(03)00017-X

Köster, E. P. (2009). Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference, 20*(2), 70-82. doi:http://dx.doi.org/10.1016/j.foodqual.2007.11.002

Köster, E. P., Couronne, T., Léon, F., Lévy, C., & Marcelino, A. S. (2003). Repeatability in hedonic sensory measurement: a conceptual exploration. *Food Quality and Preference, 14*(2), 165-176. doi:https://doi.org/10.1016/S0950-3293(02)00075-7

Köster, E. P., & Mojet, J. (2007). 11 - Boredom and the reasons why some new food products fail. In H. MacFie (Ed.), *Consumer-Led Food Product Development* (pp. 262-280): Woodhead Publishing.

Köster, E. P., & Mojet, J. (2015). From mood to food and from food to mood: A psychological perspective on the measurement of food-related emotions in consumer research. *Food Research International, 76*, 180-191. doi:10.1016/j.foodres.2015.04.006

Kostyra, E., Wasiak-Zys, G., Rambuszek, M., & Waszkiewicz-Robak, B. (2016). Determining the sensory characteristics, associated emotions and degree of liking of the visual attributes of smoked ham. A multifaceted study. *LWT - Food Science and Technology, 65*, 246-253. doi:https://doi.org/10.1016/j.lwt.2015.08.008

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421. doi:https://doi.org/10.1016/j.biopsycho.2010.03.010

Lagast, S., Gellynck, X., Schouteten, J. J., De Herdt, V., & De Steur, H. (2017). Consumers' emotions elicited by food: A systematic review of explicit and implicit methods. *Trends in Food Science & Technology, 69*, 172-189. doi:https://doi.org/10.1016/j.tifs.2017.09.006

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In *Emotion.* (pp. 25-59). Thousand Oaks, CA, US: Sage Publications, Inc.

Lawless, H. T., & Heymann, H. (1999). *Sensory Evaluation of Food*. Gaithersburg, Maryland: Aspen Publishers, Inc.

Lazarus, R. S. (1981). A cognitivist's reply to Zajonc on emotion and cognition. *36*(2), 222-223. doi:10.1037/0003-066X.36.2.222

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist, 37*(9), 1019-1024. doi:10.1037/0003-066X.37.9.1019

Liu, R., Hannum, M., & Simons, C. T. (2019). Using immersive technologies to explore the effects of congruent and incongruent contextual cues on context recall, product evaluation time, and preference and liking during consumer hedonic testing. *Food Research International, 117*, 19-29. doi:https://doi.org/10.1016/j.foodres.2018.04.024

Macht, M. (2008). How emotions affect eating: A five-way model. *Appetite, 50*(1), 1-11. doi:http://dx.doi.org/10.1016/j.appet.2007.07.002

Macht, M., & Dettmer, D. (2006). Everyday mood and emotions after eating a chocolate bar or an apple. *Appetite, 46*(3), 332-336. doi:http://dx.doi.org/10.1016/j.appet.2006.01.014

Manzocco, L., Rumignani, A., & Lagazio, C. (2013). Emotional response to fruit salads with different visual quality. *Food Quality and Preference, 28*(1), 17-22. doi:http://dx.doi.org/10.1016/j.foodqual.2012.08.014

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2016). *Sensory Evaluation Techniques* (Fifth ed.): CRC Press.

Meilgaard, M. C., Dalgliesh, C. E., & Clapperton, J. F. (1979). BEER FLAVOUR TERMINOLOGY1. *Journal of the Institute of Brewing, 85*(1), 38-42. doi:10.1002/j.2050-0416.1979.tb06826.x

Meiselman, H. L. (1996). The contextual basis for food acceptance, food choice and food intake: the food, the situation and the individual. In H. L. Meiselman & H. J. H. MacFie (Eds.), *Food choice acceptance and consumption*. London: Chapman & Hall.

Meiselman, H. L. (2013). The future in sensory/consumer research: …….....evolving to a better science. *Food Quality and Preference, 27*(2), 208-214. doi:https://doi.org/10.1016/j.foodqual.2012.03.002

Meiselman, H. L. (2015). A review of the current state of emotion research in product development. *Food Research International, 76, Part 2*, 192-199. doi:http://dx.doi.org/10.1016/j.foodres.2015.04.015

Meiselman, H. L., Johnson, J. L., Reeve, W., & Crouch, J. E. (2000). Demonstrations of the influence of the eating environment on food acceptance. *Appetite, 35*(3), 231-237. doi:10.1006/appe.2000.0360

Mojet, J., Dürrschmid, K., Danner, L., Jöchl, M., Heiniö, R.-L., Holthuysen, N. T. E., & Köster, E. P. (2015). Are implicit emotion measurements evoked by food unrelated to liking? *Food Research International, 76, Part 2*, 224-232. doi:http://dx.doi.org/10.1016/j.foodres.2015.06.031

Mora, M., Giussani, B., Pagliarini, E., & Chaya, C. (2019). Improvement of an emotional lexicon for the evaluation of beers. *Food Quality and Preference, 71*, 158-162. doi:https://doi.org/10.1016/j.foodqual.2018.06.007

Nestrud, M. A., Meiselman, H. L., King, S. C., Lesher, L. L., & Cardello, A. V. (2016). Development of EsSense25, a shorter version of the EsSense Profile®. *Food Quality and Preference, 48, Part A*, 107-117. doi:http://dx.doi.org/10.1016/j.foodqual.2015.08.005

Ng, M., Chaya, C., & Hort, J. (2013a). Beyond liking: Comparing the measurement of emotional response using EsSense Profile and consumer defined check-all-that-apply methodologies. *Food Quality and Preference, 28*(1), 193-205. doi:10.1016/j.foodqual.2012.08.012

Ng, M., Chaya, C., & Hort, J. (2013b). The influence of sensory and packaging cues on both liking and emotional, abstract and functional conceptualisations. *Food Quality and Preference, 29*(2), 146-156. doi:http://dx.doi.org/10.1016/j.foodqual.2013.03.006

Ng, M., & Hort, J. (2015). 4 - Insights into measuring emotional response in sensory and consumer research. In *Rapid Sensory Profiling Techniques* (pp. 71-90): Woodhead Publishing.

Nijman, M., James, S., Dehrmann, F., Smart, K., Ford, R., & Hort, J. (2019). The effect of consumption context on consumer hedonics, emotional response and beer choice. *Food Quality and Preference, 74*, 59-71. doi:https://doi.org/10.1016/j.foodqual.2019.01.011

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331. doi:10.1037/0033-295X.97.3.315

Paris, M. M., Carter, B. L., Traylor, A. C., Bordnick, P. S., Day, S. X., Armsworth, M. W., & Cinciripini, P. M. (2011). Cue reactivity in virtual reality: The role of context. *Addictive Behaviors, 36*(7), 696-699. doi:https://doi.org/10.1016/j.addbeh.2011.01.029

Patel, K. A., & Schlundt, D. G. (2001). Impact of moods and social context on eating behavior. *Appetite, 36*(2), 111-118. doi:http://dx.doi.org/10.1006/appe.2000.0385

Petit, C., & Sieffermann, J. M. (2007). Testing consumer preferences for iced-coffee: Does the drinking environment have any influence? *Food Quality and Preference, 18*(1), 161-172. doi:https://doi.org/10.1016/j.foodqual.2006.05.008

Phaf, R. H., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology, 5*(378). doi:10.3389/fpsyg.2014.00378

Pichon, A. M., Coppin, G., Cayeux, I., Porcherot, C., Sander, D., & Delplanque, S. (2015). Sensitivity of Physiological Emotional Measures to Odors Depends on the Product and the Pleasantness Ranges Used. *Frontiers in Psychology, 6*(1821). doi:10.3389/fpsyg.2015.01821

Piqueras-Fiszman, B., Alcaide, J., Roura, E., & Spence, C. (2012). Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it. *Food Quality and Preference, 24*(1), 205-208. doi:http://dx.doi.org/10.1016/j.foodqual.2011.08.011

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. doi:http://dx.doi.org/10.1016/j.foodqual.2014.01.007

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. doi:10.1016/j.foodqual.2013.09.002

Piqueras-Fiszman, B., & Jaeger, S. R. (2014c). The impact of the means of context evocation on consumers' emotion associations towards eating occasions. *Food Quality and Preference, 37*, 61-70. doi:10.1016/j.foodqual.2014.04.017

Piqueras-Fiszman, B., & Jaeger, S. R. (2015). The effect of product–context appropriateness on emotion associations in evoked eating occasions. *Food Quality and Preference, 40*, 49-60. doi:http://dx.doi.org/10.1016/j.foodqual.2014.08.008

Piqueras-Fiszman, B., & Jaeger, S. R. (2016). Consumer segmentation as a means to investigate emotional associations to meals. *Appetite, 105*, 249-258. doi:10.1016/j.appet.2016.05.034

Piqueras-Fiszman, B., Kraus, A. A., & Spence, C. (2014). "Yummy" versus "Yucky"! Explicit and implicit approach–avoidance motivations towards appealing and disgusting foods. *Appetite, 78*, 193-202. doi:https://doi.org/10.1016/j.appet.2014.03.029

Pober, J. (2018). What Emotions Really Are (In the Theory of Constructed Emotions). *Philosophy of Science, 85*(4), 640-659. doi:10.1086/699158

Porcherot, C., Petit, E., Giboreau, A., Gaudreau, N., & Cayeux, I. (2015). Measurement of self-reported affective feelings when an aperitif is consumed in an ecological setting. *Food Quality and Preference, 39*, 277-284. doi:https://doi.org/10.1016/j.foodqual.2014.07.016

Prescott, J. (2017). Some considerations in the measurement of emotions in sensory and consumer research. *Food Quality and Preference, 62*, 360-368. doi:https://doi.org/10.1016/j.foodqual.2017.04.005

Resurreccion, A. V. A. (1998). *Consumer Sensory Testing For Product Development*. Gaithersburg: Aspen Publishers.

Rozin, P., Markwith, M., & Ross, B. (1990). The sympathetic magical law of similarity, nominal realism and neglect of negatives in response to negative labels. *Psychological Science, 1*(6), 383-384. doi:DOI 10.1111/j.1467-9280.1990.tb00246.x

Rozin, P., & Tuorila, H. (1993). Simultaneous and temporal contextual influences on food acceptance. *Food Quality and Preference, 4*(1), 11-20. doi:http://dx.doi.org/10.1016/0950-3293(93)90309-T

Russel, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. doi:10.1037/h0077714

Samant, S. S., Chapko, M. J., & Seo, H.-S. (2017). Predicting consumer liking and preference based on emotional responses and sensory perception: A study with basic taste solutions. *Food Research International, 100*, 325-334. doi:https://doi.org/10.1016/j.foodres.2017.07.021

Schmuckler, M. A. (2001). What Is Ecological Validity? A Dimensional Analysis. *Infancy, 2*(4), 419-436. doi:10.1207/s15327078in0204_02

Sester, C., Deroy, O., Sutan, A., Galia, F., Desmarchelier, J.-F., Valentin, D., & Dacremont, C. (2013). "Having a drink in a bar": An immersive approach to explore the effects of context on drink choice. *Food Quality and Preference, 28*(1), 23-31. doi:10.1016/j.foodqual.2012.07.006

Silva, A. P., Jager, G., Van Bommel, R., Van Zyl, H., Voss, H.-P., Hogg, T., . . . De Graaf, C. (2016). Functional or emotional? How Dutch and Portuguese conceptualise beer, wine and non-alcoholic beer consumption. *Food Quality and Preference, 49*, 54-65. doi:10.1016/j.foodqual.2015.11.007

Silva, A. P., Jager, G., Van Bommel, R., Van Zyl, H., Voss, H. P., Pintado, M., . . . De Graaf, C. (2014). Emotional and functional conceptualisations for beer, wine and non-alcoholic beer consumption in the Netherlands and in Portugal. *Appetite, 83*, 350. doi:http://dx.doi.org/10.1016/j.appet.2014.06.055

Silva, A. P., Jager, G., Van Zyl, H., Voss, H.-P., Pintado, M., Hogg, T., & De Graaf, C. (2015). Cheers, Proost, Saúde: Cultural, Contextual and Psychological Factors of Wine and Beer Consumption in Portugal and in the Netherlands. *Critical Reviews in Food Science and Nutrition, 57*(7), 00-00. doi:10.1080/10408398.2014.969396

Silva, A. P., Jager, G., Voss, H.-P., Van Zyl, H., Hogg, T., Pintado, M., & De Graaf, C. (2017). What's in a name? The effect of congruent and incongruent product names on liking and emotions when consuming beer or non-alcoholic beer in a bar. *Food Quality and Preference, 55*, 58-66. doi:10.1016/j.foodqual.2016.08.008

Silva, A. P., Voss, H.-P., Van Zyl, H., Hogg, T., De Graaf, C., Pintado, M., & Jager, G. (2019). Effect of adding hop aroma in beer analysed by temporal dominance of sensations and emotions coupled with temporal liking. *Food Quality and Preference, 75*, 54-63. doi:https://doi.org/10.1016/j.foodqual.2019.02.001

Sinesio, F., Moneta, E., Porcherot, C., Abbà, S., Dreyfuss, L., Guillamet, K., . . . McEwan, J. A. (2019). Do immersive techniques help to capture consumer reality? *Food Quality and Preference, 77*, 123-134. doi:10.1016/j.foodqual.2019.05.004

Sinesio, F., Saba, A., Peparaio, M., Saggia Civitelli, E., Paoletti, F., & Moneta, E. (2018). Capturing consumer perception of vegetable freshness in a simulated real-life taste situation. *Food Research International, 105*, 764-771. doi:https://doi.org/10.1016/j.foodres.2017.11.073

Spence, C. (2013). Multisensory flavour perception. *Curr Biol, 23*(9), R365-369. doi:10.1016/j.cub.2013.01.028

Spence, C., & Van Doorn, G. (2017). Does the Shape of the Drinking Receptacle Influence Taste/Flavour Perception? A Review. *Beverages, 3*.

Spinelli, S., De Toffoli, A., Dinnella, C., Laureati, M., Pagliarini, E., Bendini, A., . . . Monteleone, E. (2018). Personality traits and gender influence liking and choice of food pungency. *Food Quality and Preference, 66*, 113-126. doi:https://doi.org/10.1016/j.foodqual.2018.01.014

Spinelli, S., Masi, C., Dinnella, C., Zoboli, G. P., & Monteleone, E. (2014). How does it make you feel? A new approach to measuring emotions in food product experience. *Food Quality and Preference, 37*, 109-122. doi:http://dx.doi.org/10.1016/j.foodqual.2013.11.009

Spinelli, S., Masi, C., Zoboli, G. P., Prescott, J., & Monteleone, E. (2015). Emotional responses to branded and unbranded foods. *Food Quality and Preference, 42*, 1-11. doi:http://dx.doi.org/10.1016/j.foodqual.2014.12.009

Stelick, A., & Dando, R. (2018). Thinking outside the booth—the eating environment, context and ecological validity in sensory and consumer research. *Current Opinion in Food Science, 21*, 26-31. doi:https://doi.org/10.1016/j.cofs.2018.05.005

Stelick, A., Penano, A. G., Riak, A. C., & Dando, R. (2018). Dynamic Context Sensory Testing–A Proof of Concept Study Bringing Virtual Reality to the Sensory Booth. *Journal of Food Science, 83*(8), 2047-2051. doi:10.1111/1750-3841.14275

Stone, H., Bleibaum, R. N., & Thomas, H. A. (2012). *Sensory Evaluation Practices* (4th ed.): Elsevier.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53-55. doi:10.5116/ijme.4dfb.8dfd

Thomson, D. M. H., & Crocker, C. (2014). Development and evaluation of measurement tools for conceptual profiling of unbranded products. *Food Quality and Preference, 33*, 1-13. doi:http://dx.doi.org/10.1016/j.foodqual.2013.10.008

Thomson, D. M. H., & Crocker, C. (2015). Application of conceptual profiling in brand, packaging and product development. *Food Quality and Preference, 40, Part B*, 343-353. doi:http://dx.doi.org/10.1016/j.foodqual.2014.04.013

Ung, C.-Y., Menozzi, M., Hartmann, C., & Siegrist, M. (2018). Innovations in consumer research: The virtual food buffet. *Food Quality and Preference, 63*, 12-17. doi:https://doi.org/10.1016/j.foodqual.2017.07.007

Van Herpen, E., Van den Broek, E., Van Trijp, H. C. M., & Yu, T. (2016). Can a virtual supermarket bring realism into the lab? Comparing shopping behavior using virtual and pictorial store representations to behavior in a physical store. *Appetite, 107*, 196-207. doi:https://doi.org/10.1016/j.appet.2016.07.033

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 countries. *Food Quality and Preference, 41*, 201-213. doi:https://doi.org/10.1016/j.foodqual.2014.12.003

Van Zyl, H., & Meiselman, H. L. (2016). An update on the roles of culture and language in designing emotion lists: English, Spanish and Portuguese. *Food Quality and Preference, 51*, 72-76. doi:https://doi.org/10.1016/j.foodqual.2016.02.019

Varela, P., & Ares, G. (2014). *Novel Techniques in Sensory Characterization and Consumer Profiling*: CRC Press.

Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference, 49*, 119-128. doi:http://dx.doi.org/10.1016/j.foodqual.2015.12.002

Wakeling, I. N., & MacFie, H. J. H. (1995). Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples

from t may be tested. *Food Quality and Preference, 6*(4), 299-308. doi:https://doi.org/10.1016/0950-3293(95)00032-1

Wan, X., Zhou, X., Woods, A. T., & Spence, C. (2015). Influence of the glassware on the perception of alcoholic drinks. *Food Quality and Preference, 44*, 101-110. doi:http://dx.doi.org/10.1016/j.foodqual.2015.03.018

Weber, A. J., King, S. C., & Meiselman, H. L. (2004). Effects of social interaction, physical environment and food choice freedom on consumption in a meal-testing environment. *Appetite, 42*(1), 115-118. doi:https://doi.org/10.1016/j.appet.2003.10.001

Yang, Q., Dorado, R., Chaya, C., & Hort, J. (2018). The impact of PROP and thermal taster status on the emotional response to beer. *Food Quality and Preference, 68*, 420-430. doi:10.1016/j.foodqual.2018.03.001

Yen, M., & Lo, L. H. (2002). Examining test-retest reliability: an intra-class correlation approach. *Nurs Res, 51*(1), 59-62. doi:10.1097/00006199-200201000-00009

Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77*(3), 600-619. doi:10.1037/0022-3514.77.3.600

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151-175. doi:10.1037/0003-066X.35.2.151