The ability of explicit and implicit measures of emotional response to discriminate milk and yoghurt products

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Food Technology

at Massey University, Palmerston North,

New Zealand

Rachel Taylor

2021

ABSTRACT

Over the last decade there has been growing interest in measuring emotional response to foods alongside hedonic liking to better understand consumer choice, and ultimately predict purchasing behaviour. Whilst the measurement of emotional response has become more common, there is a lack of consensus regarding the appropriate methods to record emotional response, with some researchers utilising explicit measures such as questionnaires and others using implicit measures including facial expressions, skin temperature, skin conductance and heart rate. This study aimed to assess the ability of select implicit and explicit methods of measuring emotional response and liking to differentiate between products, the sensitivity of these methods to small differences in sensory characteristics, and the effect of changing the consumption context on emotional response measured using these methods.

Here, participants (n = 60) tasted milk and yoghurt samples across two sessions, one with no context and one where the participants imagined a scenario relevant to when they would consume milk or yoghurt. Implicit emotional response was measured using electrodermal activity and by recording facial expressions using two methods; measuring the movements of the *corrugator supercilli, zygomaticus major and levator labii superioris* muscles using facial electromyography and using facial expression analysis software on videos of each participant's face during the product evaluation. Explicit emotional response was recorded using a RATA variant of the EsSense 25 profile and hedonic liking was also recorded.

Hedonic liking and select EsSense 25 lexicon terms were found to discriminate products within the milk and yoghurt categories, however the patterns of liking and self-reported emotional response were dependent on participant. For all lexicon terms there was a cluster of participants who were not emotionally engaged, although the size of this cluster varied. Low emotional engagement was also seen for facial EMG, with each muscle having a cluster of participants where there was little difference in muscle activity between the products. Despite this, *corrugator* and *levator* muscle activity were able to differentiate the disliked milk and yoghurt samples, and *zygomaticus* activity was able to discriminate the most liked yoghurt sample. However, more research is needed to determine the ability to measure emotional response through facial expressions using facial EMG and also for FEA software as no meaningful data was able to be extracted using this method.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Professor Joanne Hort, Dr Amanda Dupas de Matos, and Dr Peter Cannon for their invaluable guidance, patience, and support throughout this project.

Thank you to Fonterra for funding this project, and particular thanks to the Sensory Team at Fonterra Research and Development Centre for providing the milk and yoghurt products and support throughout testing.

I would also like to express my gratitude to Malcolm Loudon for creating the stimulus presentation software for this project, and the wider Psychology Department at Massey University for allowing me to take over one of their laboratories for three months.

Also, thank you to the team at iMotions for teaching me how to use their software and their support during the extraction and processing of the data.

Finally, thank you to my family and friends for their unwavering support and encouragement throughout this past year.

Ethics approval for this research as described in this thesis was obtained from the Massey University Human Ethics Committee: Southern A, Application 20/30.

Table of Contents

List of Tables

LIST OF FIGURES

PREFACE

Sensory science seeks to understand how consumers interact with products with the aim of predicting consumer choice. Consumer response to products is generally measured with hedonic liking ratings (De Beuckelaer, Zeeman, and Van Trijp 2015), however there has been a growing trend in the sensory field with many researchers incorporating emotional response into their studies. Often this is conducted via questionnaires asking participants to rate their emotions against a list of terms (explicit methods), however emotions are considered to comprise of conscious and subconscious aspects (Berridge 2018), and the methods that require participants to conceptualise what they experience may not give a complete view of the response. Implicit emotion measurements that measure the subconscious aspects of emotion, such as measuring facial expressions and autonomic nervous system responses have become more widely used, however the capabilities of these methods are not fully known.

This thesis investigates the ability of three implicit measures of emotional response (facial EMG, EDA and facial expression analysis software) to differentiate products with small differences in sensory characteristics and compares this to self-reported emotional response and liking. The thesis is separated into five sections, the first of which is a literature review covering how emotions are defined and measured, background on self-report measures of emotions and the current literature for the selected implicit methods. The next section covers the methodology used in this investigation, including product selection, piloting of temperature storage and experimental setup, and the procedures for data collection. The remaining sections cover the results from the data collection, a discussion of these results and the methods in general, and conclusions and future perspectives for this work and future investigation in this space.

1 INTRODUCTION

Understanding and predicting consumer choice is essential to the development, acceptance, and market success of new products, however gaining these insights is challenging. Generally, hedonic liking is used as a predictor, (De Beuckelaer, Zeeman, and Van Trijp 2015), however this is a single measurement which may not reflect the complexity of the consumer experience. Increasingly, researchers are including measures of emotional response alongside hedonic liking, with studies finding that there is only a partial relationship between them, (Gutjar et al. 2015). This indicates that the aspects of consumer response which can be quantified by measuring emotions are different to those which can be quantified by hedonic liking, and that these methods could be utilised in tandem to provide greater insights into the consumer experience.

1.1 DEFINING EMOTION

Emotions are a phenomenon where a mechanism based on instinct and memory elicits a response to a stimulus (Sander 2013), with the response producing autonomic nervous system (ANS) changes, facial expressions and feelings (Coppin and Sander 2016). Past experiences with similar stimuli affect the emotions elicited (Barrett 2016), and these emotions affect interactions with the stimulus (Ruth, Brunel, and Otnes 2002). It is important to note that 'emotion' and 'mood' are often considered to be different, with emotion being the rapid, short-lived response and mood longer lasting and more consciously experienced (Köster and Mojet 2015). Because of its influence on behaviour, emotional response is increasingly of interest in efforts to understand consumer response to foods and what drives purchasing behaviour. Typically, emotional response to foods is measured using lexicons of emotion terms where participants rate the intensity of each emotion using, for example, the EsSense Profile®. However, these methods only measure the aspects of emotion that participants can conceptualise and therefore additional insights may be gained if behavioural (such as facial expression) and physiological (ANS) aspects are also measured.

1.2 EXPLICIT EMOTIONAL RESPONSE

Psychology studies have asked participants to rate their emotional response for many decades, using either predefined emotional lexicons such as Positive Affect and Negative Affect Scale (Watson 1988), or situation-specific lexicons created with the input of the participants (Richins 1997). Similar methods have been used in studies investigating emotional response to foods, with general-purpose lexicons including the EsSense Profile® (King, Meiselman, and Carr 2010), the shortened version called EsSense 25 (Nestrud et al. 2016), and emotion lexicons developed for specific product categories such as beer (Beyts et al. 2017) and coffee (Bhumiratana, Adhikari, and Chambers 2014). Typically, consumers are asked to rate the intensity that they felt each emotion using a 5-point scale (Larosa et al. 2021, Samant and Seo 2020), or occasionally a 10cm line scale (Mora, Urdaneta, and Chaya 2018, Mora, Urdaneta, and Chaya 2019). However, rating 39 terms (EsSense Profile®) or 25 (EsSense 25) for multiple samples may cause participants to experience fatigue or boredom. Some studies use check-all-that-apply (CATA) versions of these lexicons to reduce the time and difficulty of the task (Jaeger et al. 2018, Kong et al. 2020), however reducing the rating of emotion intensity to a binary response may decrease the discrimination ability of the emotions measured (King, Meiselman, and Carr 2010). Rate-all-that-apply (RATA) allows consumers to ignore terms that are not relevant but give intensity ratings for those that are, (Ares et al. 2014, Low et al. 2021, Ng, Chaya, and Hort 2013). In addition to emotion lexicons, new methodologies have been created from using emoji lexicons (Jaeger et al. 2017) to displaying emotion terms and sensory characteristics in a wheel format (Schouteten et al. 2015).

The EsSense Profile® and EsSense 25 profiles are often used as a comparison for testing new methods of explicit emotional response (Kanjanakorn and Lee 2017, Spinelli et al. 2014), and also as a comparison for implicit measures of emotional response such as facial expression analysis (Leitch et al. 2015, Mehta et al. 2021) and electrodermal activity (Samant, Chapko, and Seo 2017, Samant and Seo 2020, Samant and Seo 2019).

This review summarises the available literature concerning FEA, EMG, and EDA in a food sensory application. The findings, study design, benefits and limitations of the methods are also discussed, with the aims of assessing the strengths and limitations of each method for use in measuring emotional response in product testing, investigating the application of context in studies using these methods, and identifying gaps in the literature for future research directions.

1.3 BEHAVIOURAL AND IMPLICIT EMOTIONAL RESPONSE

A specific objective of this Master's was to review the use of three selected behavioural and physiological methods: facial expression analysis (FEA), facial electromyography (EMG), and electrodermal activity (EDA); and to assess their relative limits and benefits for application in measuring and assessing consumer emotional response in food product testing. This review forms a substantial part of this chapter.

1.3.1 Facial Expression Analysis

Humans have been interested discerning emotion and intention from facial expressions of others for centuries. In 1862, Duchenne published the first scientific paper on the movement of facial muscles of humans to form facial expressions (Duchenne de Boulogne 1862), and 10 years later, Charles Darwin published 'The Expression of the Emotions in Man and Animals' (Darwin 1872). Researchers have continued to be interested in reading the emotions of people and animals around us, with recently published papers exploring the use facial expressions to assess emotional state (Dolensek et al. 2020), or emotional response (van Bommel et al. 2020). The Facial Action Coding System (FACS) was published in 1978 and gave instructions on how to quantify the movements of facial muscles recorded in videos using 46 specific movements called ‘action units’ (Ekman, Friesen, and Hager 1978). Analysing the emotions of participants using FACS is time consuming with one minute of video requiring an hour to code by a trained individual (Donato et al. 1999). Whilst other methods of manual coding have been used to measure emotional response such as the Facial Expression Coding System (FACES) (Kring and Sloan 2007), automated coding systems are becoming more widely used. These systems use models that identify the face and recognise facial landmarks, then measure the movement of the action units and categorise them into emotional and valence/arousal responses (Martinez et al. 2019). The automation of facial expression analysis allows the data analysis to be more objective as it does not depend on an individual coder’s opinion and allows for facial expression analysis to be more accessible to researchers.

Facial expression analysis is a technique that has been used in the form of manually coded videos for decades, however it is relatively new to the sensory analysis of foods field. The advent of technology has meant that most studies using food as stimuli utilise automated coding systems, but two studies have recently utilised manual coding for measuring emotional response to tasted food samples (Le Goff and Delarue 2017, Ahn and Picard 2014). The use of manual coding is explained by the Ahn and Picard study as occurring before automated facial expression analysis was commonly used in published work, and the Le Goff and Delarue study was a student project likely with a limited budget that may not have covered the cost of a software license.

Ahn and Picard (2014) used manual coding of facial valence (probably because it was the standard approach and software applications were not advanced enough at the time) in a study investigating the use of a combination of methods to predict consumer purchasing behaviour. In order to assess whether the methods could accurately predict the outcomes of the samples, two commercial soda beverages were used, one already successful in the market, and one which had failed. The experiment involved two “machines” one of which gave a 70% chance of tasting one product and 30% chance of the other, with the other machine the reverse. 39 participants each did 30 ‘trials’ which required them to choose a machine, taste the sample it selected, and then rate their liking of the sample each time and of the machines after every 5 trials. In addition to the liking measures, facial expressions and sip size were recorded throughout the experiment. In terms of liking, the two samples were deemed similar. Whilst this indicated that the participants had similar responses to both samples, the facial expression analysis found that there were significant differences in response to the samples, with the product that failed evoking more negative expressions than the successful product.

In the Le Goff and Delarue (2017) study, 100 participants tasted a total of four potato chip samples that were enriched with either insect or non-insect protein, and flavoured with two normal (chicken and barbecue) and two incongruent flavours (strawberry and blackcurrant). Videos were recorded whilst participants consumed the samples for later analysis using the FACES manual coding system (Kring and Sloan 2007). After each sample, participants rated their hedonic liking and completed an International Positive and Negative Affect Schedule Short-Form (I-PANAS-SF) questionnaire which involves rating the intensity of positive and negative moods. Interestingly, the I-PANAS-SF results were significantly more positive for participants in the insect group due to higher 'active' scores, however the negative scores were not significantly different between conditions. There were significant differences between the samples in liking, with the incongruent flavours (strawberry and blackcurrant) less liked than the normal flavours, however there was no significant difference in liking between the insect and non-insect conditions. Unsurprisingly, the incongruent flavours evoked negative facial expressions that were significantly more intense than for congruent flavours and could be distinguished from the congruent flavours using facial expressions, with similar accuracy to liking.

Facial expression analysis is a relatively new method for measuring emotional response to foods, therefore it is not unexpected that some studies have investigated this method using basic taste solutions that are known to give certain responses. Crist et al. (2018) had 46 participants taste four bitter solutions of increasing intensity whilst facial expressions were recorded, then rate their liking and perceived intensity of bitterness for each sample. As expected, there were significant differences in the liking scores between samples, with liking decreasing as the intensity of the bitterness increased. Facial expressions of disgust were positively correlated with the concentration of the solutions; however, this was less significant than the relationship between liking and bitterness. Interestingly, facial expressions that were identified as "happy" by automatic facial expression analysis were also positively correlated with the concentration of the solutions. However, these expressions were negatively correlated with liking which indicates that it was not likely to be happiness that was being recorded.

Zhi, Cao, and Cao (2017) used basic taste solutions (sourness, sweetness, bitterness, saltiness, and umami) at three or six different concentrations, evaluated over 10 sessions. Participants were asked to rate their hedonic liking of each sample, and videos were recorded during consumption and analysed for facial emotions. Of the total 50 participants, 11 had video recordings that did not meet the requirements for reasons such as facing in the wrong direction, leaving data from 39 participants. Emotions measured by facial expressions could differentiate between different tastes, although not between sweetness and water. Some facial expressions could distinguish between different concentrations of the taste solutions, and “disgust” was found to effectively distinguish between the different tastes.

Samant, Chapko, and Seo (2017) also used basic taste solutions, where 102 participants tasted four taste solutions (sweet, salty, sour, and bitter) at both ‘low’ and ‘high’ concentrations with water as a control. Alongside videos for facial expression analysis, EDA, heart rate (HR), skin temperature (ST) were measured, and participants were asked to rate their liking of each sample, its perceived intensity, and their emotional response to (EsSense 25) each sample. Following this, the participants had a short break and then re-tasted the samples in another room to rank them based on their preference. Results showed that higher levels of “surprise” and “joy” emotions and lower levels of negative facial emotions such as “disgust”, “fear” were associated with samples with greater liking scores. Samples with a higher preference rank were associated with lower levels of negative facial emotions such as “disgust” and “fear” and higher levels of “surprise” and “joy”. Facial expression analysis could discriminate between samples, however, EsSense 25 was more discriminating. This study was particularly interested in building models to accurately predict liking and preference rank and found that a combination of methods gave a more accurate model than one method alone though the R^2 values of these models were low (0.50 and 0.10 respectively).

In a study where basic tastes were added to chocolates, no significant difference between the samples was detected through participant emotional response measured using facial expression analysis (Gunaratne et al. 2019). In this, 45 participants tasted five samples of 70% cocoa chocolate with nothing, sugar, salt, citric acid, and monosodium glutamate added to represent bitter, sweet, salty, sour, and umami respectively. The samples were tasted in a sensory booth whilst participant ST, HR, and facial expressions were recorded; participants also rated their overall liking and Check-All-That-Apply (CATA) emotional response to each of the samples using a chocolate relevant lexicon. Whilst there were significant differences in the perceived intensities of all the basic tastes and in the liking and some emotional scores, there were no significant differences between samples based on the emotional response measured by facial expressions.

Another study used four chocolate samples and 15 images as the stimulus to assess cross-cultural differences in emotional response (Torrico et al. 2018). 60 participants (40% Western and 60% Asian) had their ST, HR and facial expressions measured and rated their emotional response on a 3-point ‘face scale’. Hedonic liking was also rated for each of the food samples. From facial expression analysis, results indicated there were no significant differences between the emotions elicited by each of the images, and “neutral” was the emotion expressed with the highest intensity for both cultural groups. For the chocolate samples the facial expression analysis gave no significant differences in emotional response between the chocolate samples nor between the cultural group. However, this is not surprising as there were no significant differences found in liking and emotion face scale response between samples or cultural groups.

Garcia-Burgos and Zamora (2013) used a chocolate liquor (selected for its strong bitter taste) and a grapefruit juice sample to assess participant emotional response to bitterness. In this, 40 participants tasted each sample and rated their hedonic liking and wanting, and their facial expressions were recorded in the 10 seconds post-consumption. Facial expression analysis found that samples were able to be discriminated by expressions of 'disgust', 'anger', and 'neutral', however liking had greater discrimination ability than all of these.

It is important to note that finding significant differences in liking between samples does not necessarily mean that there will be a significant difference in emotional response measured by facial expression analysis. In fact, this was shown by Gonzalez Viejo et al. (2019) who investigated consumer response to beer. In this study, nine different samples of beer from three different categories were tasted by 30 consumers whilst facial expressions, infra-red thermal imagery, Electroencephalography (EEG), overall liking and sensory perception were recorded. Similar to Gunaratne et al. (2019), there were significant differences between samples in sensory perception and liking, but no significant difference between the beers detected by facial expression analysis. It should be remembered that liking and emotional response are not necessarily the same and different techniques are designed to measure different aspects of affect.

Another study that did not find significant differences between similar samples using facial expression analysis was Leitch et al. (2015). This study investigated the effect of using different sweeteners in iced tea and involved 31 participants tasting six samples (two artificial or two natural and two sucrose sweeteners) across two days. Facial expressions were measured, and participants rated their hedonic liking, and then later re-tasted the samples and rated their CATA emotional response (Modified EsSense). The authors found that liking and explicit emotional response could both distinguish one of the natural sweeteners (honey) from the other sweeteners, however facial expression analysis did not.

Similarly, a study investigating emotional response to commercial energy drinks was unable to find significant differences between the samples using facial expression analysis, (Mehta et al. 2021). In this, 30 participants tasted the two samples whilst their facial expressions were recorded using video and rated their hedonic liking and explicit emotional response using CATA on a shortened version of the EsSense profile with 21 of 39 terms. No significant differences in emotions measured with facial expressions were found, however there were significant differences in overall liking and the ratings of Active and Interested from the EsSense profile.

Kaneko et al. (2019) investigated the ability of facial expression analysis to discriminate between eight beverages (including milk, yoghurt, orange juice, black tea) and an unpleasant stimulus (diluted vinegar). 70 participants tasted the samples whilst their facial expressions, EDA, EEG, pupil diameter, and sip size responses were recorded, and then completed a self-assessment manikin (SAM). The self-reported valence and arousal of the vinegar solutions were significantly different from the beverages and they could also discriminate between the beverages. Interestingly, the emotion "disgust" from facial expression analysis was able to discriminate between the vinegar and normal beverages, however, there was no significant differences between the beverages. This implied that facial expressions can discriminate between samples that are vastly different in valence and arousal, but not necessarily between those that are more similar.

Alvarez-Pato et al. (2020) also selected samples with large hedonic differences, with participants evaluating odours such as mint and vinegar, and tasting gelatine gums with flavours such as clam and strawberry. The response of 120 consumers to these stimuli was measured using a novel system of analysing facial expressions for emotions alongside EDA, heart rate and liking. The pleasant and unpleasant stimuli were able to be differentiated by liking, however none of the implicit measures were able to differentiate between the samples.

In contrast, other studies using beverages that are more similar in terms of flavour and composition have differentiated samples using facial expression analysis. Samant and Seo (2019) used five different vegetable juices that were found to have significantly different sensory characteristics, and measured facial expressions, autonomic nervous system (ANS) responses, and explicit emotional responses (EsSense25) from 100 participants. It was found that facial expressions of "surprise" before and directly after consumption were associated with samples rated with higher liking scores, as were lower levels of facial expressions of "sadness" and "disgust" post-consumption. Whilst facial expressions had a relationship with liking, the best model to predict liking used a combination of sensory attribute scores, EsSense25 scores, and facial expressions, which highlight the importance of combining methods. In a further study using vegetable juice, Samant and Seo (2020) used facial expression analysis, ANS responses, and self-report emotional responses (EsSense25) alongside purchase intent. It was found that of the emotions from facial expression analysis, "joy" could discriminate products when consumers viewed the samples, and "fear", "contempt", "disgust" and "sadness" could discriminate after the samples were tasted. In addition, "surprise", "disgust" and "sadness" were found to contribute to purchase intent, with a positive relationship between increased levels of "surprise" and purchase intent and negative for "disgust" and "sadness".

Zhi et al. (2018) also investigated consumer emotional response to juice products using facial expression analysis alongside hedonic liking. In this study, 46 participants tasted five different samples of fruit juice whilst their facial expressions in the 5 seconds after the cup was removed from the mouth were recorded. Significant differences were found between the hedonic liking of the juice samples, meaning that the samples could be discriminated by liking. Interestingly, no one expression could discriminate between the samples, however when all seven emotions were considered there were significant differences between samples. For samples with different hedonic liking scores, the intensity of "sadness", "anger" and "surprise" expressions over time were found to discriminate samples.

Mojet et al. (2015) measured the implicit emotional response of three groups of participants (96 in total) to three pairs of yoghurts, each from a different brand. Participants tasted the samples whilst their facial expressions were recorded, rated their liking, then completed an 'emotion projection task'. In addition to tasting the samples, participants viewed images and rated their expected liking. Interestingly, whilst there were no significant differences in liking between the samples within each pair of yoghurts, some emotion terms in the emotion projection task revealed significant differences between samples in one or two (but not all three) of the pairs. Unfortunately, the results of the facial expression analysis were not reported on due to technical issues with the data collection.

In an early study on the use of facial expression analysis with tasted samples, 19 participants tasted five different flavours of commercial breakfast drinks whilst their facial expressions, HR, skin conductance, and ST were recorded (De Wijk et al. 2014). After tasting each sample, the participants rated their perceived intensity and liking. Interestingly, the liking showed no significant differences between samples, and therefore could not discriminate between the samples. However, there was a significant effect of sample on "happiness" recorded by facial expression analysis, showing some ability of happiness to discriminate, but there were no other significant relationships between sample and emotion.

In another study with small differences between the composition of the samples, Rocha et al. (2019) investigated the emotional response of 50 regular consumers to lemon verbena tea herbal infusions using five different brands. The participants could add sugar (providing they added the same amount to every sample), and then were recorded tasting the samples, and rated their liking and explicit emotional response using a modified EsSense profile. Facial expression analysis was carried out, but the dominant emotions identified from this were then entered into a temporal dominance of facial emotions analysis. From this, 'neutral' was found to be the most dominant facial expression across all the samples but was removed to make the size of the differences between the other emotions more obvious. With 'neutral' excluded the only emotions that were then dominant across all samples were "sad" and "contempt", and the least liked samples could be distinguished by the presence of a period where "contempt" was significantly dominant. It was found that there were significant differences in liking between samples, with the least liked sample distinctly different to the other samples. In terms of the modified EsSense profile, the 'premium' sample and most liked sample evoked 'adventure' emotions whilst the least liked sample received more ratings of 'bored'. The explicit and implicit emotional measures had similar discrimination ability, with an RV value of 0.704 indicating that these methods were similar to each other.

van Bommel et al. (2020) also used facial expression analysis to assess consumer emotional response over time. In this study, 56 participants tasted yoghurts with granola pieces and rated their emotional response using temporal dominance of emotions (TDE) and liking across multiple bites. Products differentiated best by liking, however "angry", "sad", "surprised", and "bored" emotions from facial expression analysis were able to differentiate the least liked samples. Despite some emotions in the facial expression analysis and TDE having similar or the same names, there was no significant overlap between the emotions recorded from these measures.

Another study that used automated facial expression analysis in an innovative approach was De Wijk et al. (2019a) who measured emotional response to stir fried meat and meat alternatives in two different contexts: in the laboratory and in participant's homes. Participants tasted the samples in 10 sessions across two weeks, alternating between locations. Issues with internet connection and camera quality for facial expressions recordings meant that only 18 of the original 32 participants had usable data for all 10 sessions. It was found that between the samples, there were significant differences in the intensities of 'sad", "surprised" and "scared" facial expressions as well as facial valence. There were also significant differences between liking scores and sensory perception between the samples. It was found that facial expressions of "surprised" had similar discrimination ability to liking, whilst" sad", "scared" and facial valence were less discriminating. Further, the location of testing and the sample was found to have a significant effect on all responses measured using facial expressions, however there was no significant effect of location on liking and sensory perception.

Whilst facial expression analysis is generally used as an implicit measure of emotional response, some studies have used it as an explicit measure. Danner et al. (2014) used six different orange juices as stimuli in two experiments; one using facial expressions as an explicit measure and one using them as an implicit measure. In the explicit task, participants were asked to taste the sample and then make a facial expression that reflected how they felt about it, whereas in the implicit task, the 78 participants were not informed that they and their facial expressions were being recorded and throughout the task. Hedonic liking was measured in both tasks, and for both there was a high correlation between liking and facial expression. The emotions from the facial expressions recorded in both tasks could discriminate between samples, however the implicit task had the greater discrimination ability of the two, likely because a larger variety of facial expressions were recorded.

In Juodeikiene et al. (2018), participants were asked to taste different bread and chocolate samples, wait 15 seconds, give a signal to the experimenter and then form a facial expression that represented their experience. The participants were also interviewed on their attitudes towards the samples, emotions evoked by the samples and their purchase intentions. Facial expressions of "angry", "sad", "happy" and "neutral" allowed for discrimination between liked, disliked and neutral samples.

Another study that used facial expression analysis as a measure of explicit emotional response was conducted on smoked ham samples that varied in pig breed and feed type (Kostyra, Rambuszek, et al. 2016). 30 participants tasted each sample and around 10 seconds after swallowing indicated to the experimenter to begin recording facial expressions and then form a facial expression that represented their liking of the sample. Results showed that the emotion present with the highest intensity and frequency was "neutral". Further, most of the variability in the emotions recorded using facial expressions was due to differences between individual participants rather than samples. This study was part of a larger investigation into the acceptability of smoked ham samples where the same participants also rated their liking of tasted samples and images, and viewed images with their eye movement tracked (Kostyra, Wasiak-Zys, et al. 2016). In this case, the emotion "surprised" had a significant correlation with liking, however, it was the only emotion to do so.

Zokaityte et al. (2020) also utilised facial expressions as an explicit measure of consumers' emotional response to nutraceutical beverages. After tasting a sample, participants were asked to signal to the researcher and then form a facial expression that was representative of their liking of the sample. It was found that there was a strong correlation between the emotion "happy" and overall liking of the samples, as well as a negative correlation between "angry" and liking.

Whilst facial expressions can be analysed as discrete emotions such as "disgust' or 'joy', valence and arousal can also be extracted using facial expression analysis. Brouwer et al. (2019) used facial expression analysis alongside EDA and other measures in a study investigating valence and arousal responses to cooking with either 'basic' or 'premium' ingredients (35 and 39 participants respectively). Facial expressions were recorded throughout the cooking and tasting process, but it was found that there was no significant difference in facial arousal or valence between the two groups of participants during any phase of cooking or tasting. However, this was not the only measure to show no significant difference between the 'basic' and 'premium' groups, as the self-reported valence and arousal scores also saw no significant difference. It is possible that the lack of significant differences between the ingredient groups could be due to each ingredient type being tested on a separate set of participants, therefore the differences in the reactions of individual participants cannot be accounted for.

1.3.2 Facial electromyography (EMG) in emotional response measurement

Another method of measuring facial movements is facial electromyography (EMG), which measures electrical impulses across muscles in the face. Unlike facial expression analysis which analyses video recordings of participants, facial EMG is measured by placing gel-filled electrodes on the surface of the participant's skin over a specific muscle group. The signals from these electrodes show the changes in the electrical potential that occur when muscles contract and relax and can detect small changes in facial muscle movements. Three muscle groups are commonly measured with facial EMG, with studies using just one or a combination. The *zygomaticus major* (zygomaticus) which is involved with smiling, *corrugator supercilii* (*corrugator*) which is involved with frowning, and *levator labii superioris* (levator) which is involved in nose-wrinkling (Nath, Cannon, and Philipp 2019). At present there is only one study that used facial EMG to measure affective response to tasted food samples, (Sato et al. 2020). This study measured the activity of *zygomaticus* and *corrugator* of participants as they tasted flavoured gels, with participants rating their liking, wanting valence and arousal for each sample. To reduce noise in the *zygomaticus* data, participants were asked to refrain from chewing the samples during the period when the EMG data was recorded. *Corrugator* activity was found to have negative associations with liking, wanting and valence, however there were no associations with explicit measures for the *zygomaticus,* potentially due to the inhibition of chewing.

Another study found in the literature used tasted samples to evaluate EMG as a measure of emotional response, however this was conducted on flavours in oral care products rather than foods (Chen et al. 2018). The study was primarily interested in arousal and valence responses measured using the Self-Assessment Manikin (SAM), however participants also rated six terms from the EsSense profile ("good", "loving", "pleasant", "disgusted", "aggressive", and "worried"). The *zygomaticus* appeared to be the only muscle measured in this study, and only 12 of the 24 participants had their muscle activity recorded, data from 10 included in the data analysis. Due to the small data set it is not surprising that there were no meaningful conclusions drawn about the potential of *zygomaticus* activity as a measure of emotional response, only that no significant relationship was found.

Due to the lack of literature concerning emotion measurement using facial EMG with tasted food samples, five other studies close to meeting the criteria were included which studied food images, aromas and personal care products. It is possible that no other EMG emotion related studies exist due to poor methodological approaches and decisions to not publish.

Soussignan et al. (2015) investigated the effect of emotional communication with virtual characters on participant response to food images. In this, participants were asked to view videos where the virtual character looked at a food image, formed a facial expression of joy, neutral or disgust and then either returned its gaze to the participant or continued looking at the food. The movements of participant *corrugator*, *levator*, and *zygomaticus* muscles were recorded using facial EMG whilst they viewed the stimuli, and then rated their liking and wanting of the food in each image. Increased *zygomaticus* activity was found to be linked to food images with higher liking scores, and inversely, increased *corrugator* activity was linked to disliked images. Interestingly, the facial expression displayed by the avatar had a significant effect on participant facial muscle movements and the liking scores indicating that these may be changed by social context. Other studies have investigated the effects of social context on facial muscle movements and hedonic liking using food images as stimuli. One such study, Nath, Cannon, and Philipp (2019), used facial EMG to measure participant movement of the *corrugator*, *levator*, and *zygomaticus* muscles whilst viewing food images either alone or observed. *Corrugator* and *levator* activity were found to be negatively correlated with liking and *zygomaticus* activity to be positively correlated with liking, although the effects were small (-0.18, -0.10 and 0.04 respectively) but of consequence. The participants displayed significantly lower *levator* activity, indicating that the presence of a stranger alters the 'disgust' facial expression. In a later study, Nath, Cannon, and Philipp (2020) further investigated the effect of social context on participant liking and facial muscle movement in response to food images. In this, participants either completed the testing alone, with a friend or with a stranger completing the testing at the same time. Participants who completed testing with a stranger rated their liking of foods significantly lower than those who participated alongside a friend. The social context also had a significant effect on the relationship between *zygomaticus* activity and liking, with the 'friends' condition having a negative relationship as opposed to the expected positive relationship seen in the 'strangers' condition. Whilst social context was not the focus of this literature review, it is important to note that who the participant is with during testing can affect perception and also the way in which their facial expressions relate to it.

The effect of context was also investigated by Sato, Yoshikawa, and Fushiki (2020) who recorded *zygomaticus* and *corrugator* activity as participants viewed food images with and without nutrition information. Participants also rated their liking, wanting, arousal and valence of each food image. It was found that *zygomaticus* activity positively correlated with liking, wanting and valence, however there were no significant correlations with the *corrugator*. Further, the inclusion of nutrition information had no significant effect on the activity of the *corrugator* and *zygomaticus* muscles.

Beyts, Chaya, Dehrmann, James, Smart, & Hort (2017), investigated emotional response to beer aromas. This study used EMG to measure the activity of two muscle groups (*corrugator* and *zygomaticus*), and asked participants to rate their hedonic liking and explicit emotional response using EsSense 25. *Corrugator* activity was found to be able to differentiate between beer samples better than *zygomaticus* activity, however hedonic liking and the EsSense 25 lexicon were more discriminating than the activity of either muscle.

1.3.3 Electrodermal activity (EDA) in emotional response measurement

Electrodermal activity (EDA) measures an aspect of the autonomic nervous system. It measures neurally mediated effects on sweat gland permeability as changes in the resistance of the skin to a small electrical current which occurs when the body has an emotional response to stimuli (Kenney and Adhikari 2016). An EDA signal is composed of a slowly varying "tonic" component and a rapidly changing "phasic component" (Benedek and Kaernbach 2010), with studies choosing to report on the skin conductance response (SCR) as a whole (De Wijk et al. 2014), or the phasic component alone (Brouwer et al. 2019, Samant and Seo 2019). Some studies report both (Sargent et al. 2020). EDA is measured using two electrodes placed on two fingers of the participant's non-dominant hand (Samant, Chapko, and Seo 2017), allowing it to be used in locations outside of the laboratory (Xu et al. 2019).

In food sensory studies, EDA is often used alongside other autonomic nervous system measures such as ST and HR. In a study on breakfast drinks, De Wijk et al. (2014) used these measures alongside facial expression analysis software to investigate their ability to discriminate between samples and their relationship with liking. It was found that there was no significant effect of sample or replicate on the electrodermal activity of the participants. Similar results were found in a study investigating the use of the EsSense25 lexicon alongside EDA, HR, and facial expression analysis techniques to predict liking and preference rank of basic taste solutions (Samant, Chapko, and Seo 2017). In this latter study, the authors found that the EDA responses of participants had no significant relationship with either the liking and preference rank of the samples and made very little contribution to the models created to predict them. Another later study on vegetable juice samples using EDA and the EsSense25 lexicon to investigate participant emotional response had very similar results, with no significant relationship between EDA and liking (Samant and Seo 2019). Interestingly, there was a significant relationship between the EDA recorded when participants evaluated the aroma of the samples and the preference rank, however the R^2 value of the resulting model was very small (0.04). In a further study with vegetable juice samples, Samant and Seo (2020) found no significant difference in EDA when observing, sniffing or tasting the samples despite there being significant differences in facial expressions, sensory characteristics and purchase intent.

Another study that measured consumer response to odours and tasted samples was Alvarez-Pato et al. (2020) where EDA, heart rate, and a novel method of interpreting emotions from facial expressions were recorded alongside liking. Participants sniffed five odour samples and tasted gelatine gums with five flavours, (three pleasant and two unpleasant). It was found that there were no strong correlations between EDA and liking or any other measure.

EDA has also been used as an implicit measure of arousal and valence aspects of emotional response alongside the Self-Assessment Manikin. Kaneko et al. (2019) measured participant EDA, pupil diameter, sip size, and facial expressions in response to eight 'accepted drinks' and diluted vinegar. The EDA of participants was able to discriminate between the vinegar and other samples however this was expected as the valence and arousal reported by the participants for the vinegar sample were significantly different to all the other samples. When the differences between samples was more subtle, such as between the 'accepted drinks' the EDA measures were unable to discriminate between the samples, despite explicit valence and arousal demonstrating this capability.

Another study using EDA alongside SAM as a measure of arousal and valence investigated consumer experience when cooking and eating a meal made with chicken or mealworms (Brouwer et al. 2017). Whilst the EDA was measured throughout the cooking and tasting process, it was only able to discriminate between the two ingredients at two points; at first exposure to the ingredient, and whilst the food was cooling. Whilst EDA could determine whether participants were cooking with very different foods (chicken and mealworms) in the work of Brouwer et al. (2017), the measure was unable to distinguish between 'premium' and 'basic' ingredients. This is consistent with what was found by Kaneko et al. (2019), where only samples with large differences in acceptability were able to be discriminated with EDA, indicating that this method on its own may not be sensitive enough to discriminate between samples that are similarly liked.

Rita, Guerreiro, and Omarji (2020) also used EDA as a measure of arousal alongside SAM in a study comparing consumer's emotional response to private label and popular brands of chocolate. Participants tasted chocolate samples from private label brands and popular brands under blind and informed conditions, with the data from 19 participants used in this study. EDA was unable to discriminate between the products under blind conditions but could when the participants were informed of the brand. Interestingly, arousal measured using SAM did not discriminate between the brands of chocolate under either condition indicating that the EDA may be measuring a different aspect of response to SAM.

Sargent et al. (2020) used EDA as a measure of participants' arousal during the preparation of hot beverages, alongside electroencephalography (EEG) measuring valence and self-report liking.

Participants were asked to complete a cognitive task before preparing and consuming a hot beverage using one of two machines, which was repeated three times. First, the participants chose the machine they used however they were required to use the other machine to make the second beverage, with the final repeat again giving the participant the freedom to choose. Both tonic and phasic measures of EDA were unable to differentiate between the machines, however there were significant interactions between the machine and choice, with differentiation between the machines during the consumption of the third beverage with the free choice of machine.

EDA has also been used to investigate the effect of context on consumer experience with products. In a study investigating the effect of location on the sensory experience and emotional response of participants consuming chocolate ice cream, Xu et al. (2019) employed EDA, SAM, and heart rate alongside temporal dominance of sensations. It was found that the eating environment had a significant effect on the skin conductance of the participants, with the largest difference in EDA between participants eating ice cream in the laboratory compared to in a university study area. As this study primarily investigated the effect of location on participant experience of ice cream, only one sample was evaluated in this study meaning that no information on the discrimination ability of the EDA measured in different locations was available.

The papers demonstrated a gap in the literature for studies using facial EMG to measure the emotional response of participants to tasted food samples. They also showed that the context of consumption is an important aspect to consider, as it may affect the data collected and subsequent results.

1.3.4 Trends in the application of EMG, FEA and EDA for emotion measurement

The limited number of studies using these techniques to assess consumer response to tasted food samples suggests that their use is still in its early phases. It is therefore not as surprising that in many of the studies discovered via this review the focus was on studying the methods, rather than investigating the specific stimuli used in the experiments. This is demonstrated by the article using basic taste solutions as the stimuli (Crist et al. 2018, Samant, Chapko, and Seo 2017, Zhi, Cao, and Cao 2017), or that incorporating basic tastes into chocolate (Gunaratne et al. 2019). Some papers also openly selected samples that were expected to give large differences in liking, affective, and emotional response, confirming that being able to find differences between similar samples was a lesser objective, as opposed to the ability to see 'a' difference using the chosen methods (Chen et al. 2018, Garcia-Burgos and Zamora 2013). Other studies used test stimuli that were unusual or were specifically designed to evoke a negative response in order to have a large difference between these and more accepted samples. One such study was Le Goff and Delarue (2017) which investigated consumer response to potato chips enriched with insect protein alongside chicken and barbeque flavours, and included strawberry and blackcurrant flavoured samples. Unsurprisingly, there was a significant difference between the normal flavours and the incongruent flavours, however unexpectedly there was no significant difference between the insect and non-insect groups. This indicates that the inclusion of the incongruent samples may have masked a more representative participant response to the insect protein. Kaneko et al. (2019) also selected samples expecting to see a difference, choosing to test diluted vinegar alongside regular drinks to investigate the sensitivity of facial expressions and EDA to similar and very different samples. The explicit and implicit methods could easily discriminate between the vinegar and the other samples, but the more similar samples could only be discriminated by explicit affective response and not implicit measures. Brouwer et al. (2017) also selected stimuli that were expected to give a large difference in affective response, with participants cooking with mealworms and chicken. As expected, there was a significant difference in both implicit and explicit affective response between the ingredient types, however when the ingredients differed only by quality, no significant differences could be seen between the implicit or explicit affective response to the cooking task (Brouwer et al. 2019).

There is a lack of consensus about whether facial expression analysis can discriminate between similar samples. Some studies have found no significant differences across emotions measured between similar beers (Gonzalez Viejo et al. 2019), sweeteners (Leitch et al. 2015) and chocolate samples (Torrico et al. 2018, Gunaratne et al. 2019). However facial expressions have been found to discriminate between lemon verbena tea brands (Rocha et al. 2019), fruit juices (Zhi et al. 2018), breakfast drinks (De Wijk et al. 2014) and stir-fried chicken and meat alternatives (De Wijk et al. 2019a). There is a clear gap in the literature concerning the ability of these techniques to distinguish consumer response to more similar samples within a food category which would be important for measuring and guiding new technical and product development.

From the literature it can be seen that EDA can discriminate very different samples (Kaneko et al. 2019, Brouwer et al. 2017), but not similar samples (Samant and Seo 2019, Kaneko et al. 2019, Brouwer et al. 2019). EDA has been found to not have a significant effect on the predictive power of models (Samant and Seo 2019, Samant, Chapko, and Seo 2017), however these studies used explicit emotional response measures whilst EDA is generally considered a measure of arousal. Therefore, it may be beneficial to measure EDA alongside emotional response measures in order to gain additional insights into the consumer experience.

Most studies using these methods used beverages as stimuli, with very few studies using solid foods. One reason for this may be to reduce the potential for 'noise' to be introduced into measures of facial movements due to chewing as its effects would be expected to be seen in facial expression analysis and EMG data. Whether this effect could be easily removed is uncertain from these studies as there is no published study using solid food stimuli with EMG, and facial expressions were generally analysed from recordings post-consumption.

Several studies noted difficulties in collecting facial expression recordings during consumption of samples, with cups or participant hands in front of their face (Zhi, Cao, and Cao 2017, Samant and Seo 2019), and poor lighting (De Wijk et al. 2019a). These problems do not affect EMG measurements as the muscle movements are measured via electrodes on the skin, however incorrectly applied electrodes or poor signals can cause participant data to be too noisy to use as noted in Chen et al. (2018). Mouth movements during eating may also interfere with the emotional signal. There is evidently a need to understand more about how interference from the sample delivery vessel can be reduced with FEA for food consumption studies, alongside potential interference from mouth movements for both techniques.

There appears to be opportunities to improve the protocols applied across the different methodologies to enable better data to be collected. For example, there were two different approaches to accounting for individual differences in facial muscle movements between participants; converting EMG signals to percentages of the maximum voluntary contractions of each participant (Nath, Cannon, and Philipp 2020, 2019), and adjusting the EMG signals by subtracting a baseline measure (Soussignan et al. 2015, Chen et al. 2018, Beyts et al. 2017). Interestingly for the studies that used baselines there were differences in how the baseline was determined, using the muscle activity directly before exposure to the stimulus (Soussignan et al. 2015, Beyts et al. 2017), or using the muscle activity when water was used as the stimulus (Chen et al. 2018). Further, the timing of testing should be considered to ensure that participants are tasting samples at a time of day that makes sense for when that food would normally be consumed, and that the different participants are tasting the samples at a similar time. For example Le Goff and Delarue (2017) collected data between the times of 9:00 and 18:00, which is not only a large time-frame, but 9:00 is an unusual time to be consuming potato chips. The data in studies surrounding the impact of social context all point to important considerations for researcher behaviour, in addition to understanding the likely contexts for the products under investigation, in future studies.

Another important issue highlighted by the review is the loss of participant data mentioned in several studies, sometimes due to participant behaviour, sometimes the set-up of the electrodes or faulty internet connections. This implies a necessity to pay attention to participant warm up activities and instructions, as well as the technical capability of the equipment operator and a general need to over recruited to cope with technical difficulties that may occur that are out of experimenter control.

It was also evident that in some papers researchers may have been ‘selecting’ or ‘adjusting’ the data analysis approach in search of a significant result, for example Rocha et al. (2019) removed the most dominant expression ‘neutral’ from the temporal dominance of facial emotions. This was done so that the smaller differences in the other emotions were more obvious. Another example of questionable data handling is only reporting selected results from the facial expression analysis. Kaneko et al. (2019) only reported on the disgust emotion, which either means the others were not measured or, more likely, it was the only emotion that showed anything significant. Regardless, in order to better understand the ability of facial expression analysis to discriminate samples, all emotions should be analysed for and reported on. Further, models to predict liking and preference rank created using these methods often have low predictability, for example the optimal models to predict liking and preference rank in Samant, Chapko, and Seo (2017) had R^2 values of 0.50 and 0.10 respectively. The R^2 value of a model shows the proportion of the variation that can be explained by the independent variables in the model, therefore in the liking model only 50% of the variation in the liking scores was explained by the model. In a later study which tested different samples and included more measures of emotional response, the optimal model to predict liking had a slightly higher R^2 value of 0.61, however the preference rank model had very little predictive ability with an R^2 value of 0.04 (Samant and Seo 2019).

Evaluating the potential for EMG to measure the emotional response to food samples is challenging from the available literature as only one study measured affective response to tasted food samples (Sato et al. 2020), and only one (Beyts et al. 2017) used explicit emotional response as a comparison. Despite this, the available studies show some promising relationships between facial muscle movements and explicit measures of liking (Soussignan et al. 2015, Nath, Cannon, and Philipp 2020, 2019), and emotional response (Beyts et al. 2017). Whilst EMG is sensitive to small movements of facial muscles, the electrodes must have good contact with the skin for the data to be useful. Both Beyts et al. (2017) and Chen et al. (2018) had to exclude data from participants who had issues with signal strength or noise in the signals.

A trend that was seen in both the EMG studies and some facial expression analysis studies was that facial movements associated with negative emotions have a stronger relationship with liking than those related to positive emotions. Nath, Cannon, and Philipp (2019) found that the negative relationship between c*orrugator supercilii* and *levator* activity had larger effects than the positive relationship between *zygomaticus major* activity and liking (-0.18 and -0.10 compared to 0.04). Similarly, De Wijk et al. (2014) found that facial expressions with negative valence (sad, scared, anger, surprise) were negatively related to liking, whereas the expected positive relationship between liking and ‘happiness’ was not present. Instead, the relationship between liking and facial expressions of ‘happiness’ was also negative, which was also seen in Crist et al. (2018). This may be due to the presence of the experimenters in the room during testing, as the relationship between facial expressions of happiness and liking has been found to be affected by social context (Nath, Cannon, and Philipp 2020). This is supported by Danner et al. (2014) who found that when asked to form a facial expression, expressions of ‘happy’ were positively correlated with liking, however when the participants were not aware they were being recorded, expressions of ‘happy’ had a small negative correlation with liking. This demonstrates that expressions of ‘happy’ likely serve a communication role, and therefore the presence of other people or the knowledge that facial movements are being recorded may alter how they are used.

Most studies in the literature found that liking gave greater discrimination between samples than implicit emotional response methods. However, Le Goff and Delarue (2017) found that whilst liking could not discriminate the insect protein enriched potato chips from the non-insect protein version, facial expression analysis could distinguish the two groups based on significant differences in positive valence facial expressions (lower for the insect group). This indicates that in this study, the explicit liking and implicit facial expression analysis are measuring different aspects of participant experience. Similarly, Ahn and Picard (2014) found that liking could not discriminate between samples, however the facial expression analysis did, showing significantly more negative facial emotions towards one of the samples. De Wijk et al. (2019a) found that whilst liking and facial expression analysis could both discriminate between samples, facial expression analysis was sensitive to the location in which the samples were consumed whereas the liking scores were not.

1.4 Considering consumption context

Consumer acceptance testing is typically conducted in a controlled environment (such as a laboratory or a central location test (CLT)) where the participant tastes the samples without any additional information. However, consumer response measured in this way may not be representative of responses in real-life situations, (King et al. 2007). There has been growing interest in investigating the effect of the context in which samples are consumed on emotional response, with several studies measuring explicit emotional response in altered contexts. Several studies have changed the location of consumption either by using in-home tests (Jaeger et al. 2020), using virtual reality (VR) to simulate realistic consumption locations (Kong et al. 2020, Worch et al. 2020), or testing both VR and a café location (Low et al. 2021). Other methods include asking participants to imagine a situation where they would consume the sample (Piqueras-Fiszman and Jaeger 2014b), and providing the participant with additional information such as viewing the packaging (Gutjar et al. 2015), calorific information (Oliveira et al. 2020), and information on the ethnic background of the sample (Kim and Hong 2020).

Implicit measures of emotion require participants to be stationary in front of a camera for facial expression analysis or be connected to recording equipment using electrodes (EDA and facial EMG), so the method of altering the consumption context must allow for this. One paper measured consumer response (using EDA) to chocolate samples with and without the knowledge of the brand, (Rita, Guerreiro, and Omarji 2020). Two papers changed the location of sample consumption; De Wijk et al. (2019a) tested in-lab and at the participants' homes (using facial expression analysis), and Xu et al. (2019) tested in a lab and in three other locations on a university campus (using EDA and heart rate). Three papers looked at the effect of social context on participants' response measured with facial EMG, two looking at the effect of the presence or absence of another person during testing (Nath, Cannon, and Philipp 2020, 2019), and one investigating the effect of emotional communication with digital avatars (Soussignan et al. 2015). Finally, two studies measured the affective responses (using EDA and a combination of EDA and facial expression analysis) of participants whilst they were cooking with different ingredients (Brouwer et al. 2019, Brouwer et al. 2017) and one study measured affective response (using EDA and EEG) whilst consumers prepared hot beverages using different machines, (Sargent et al. 2020). Although the number of studies combining implicit measures of emotion and context changes are limited, they indicate different aspects of context are an important consideration when studying emotional response.

1.5 EMOTIONAL RESPONSE TO DAIRY PRODUCTS

There are only a limited number of studies using dairy products as the stimuli, and even fewer that only use dairy products. Xu et al. (2019) used chocolate ice cream as the stimulus in their investigation into the effects of consumption context on the affective response and sensory perception, however there was only one sample used and therefore no information on the effect of location between samples was recorded. Another paper that used dairy samples measured liking and facial expressions in response to different flavours of breakfast drinks, however there was no measure of explicit emotional response (De Wijk et al. 2014). Another study measured facial expressions alongside an emotion projection task in response to 6 different yoghurt samples, however issues with data collection meant that it was not possible to report on the results of the facial expression analysis (Mojet et al. 2015). Kaneko et al. (2019) included some dairy samples (milk, buttermilk, and yoghurt) within the range of regular drinks that they tested however the different samples were not the focus of this study. This demonstrates that there is a gap in the literature concerning studies comparing dairy samples and certainly none comparing dairy samples with similar sensory characteristics using these methods.

1.6 Conclusions and opportunities

It is difficult to conclude from the literature available as to whether these methods will be successful in discriminating emotional response to dairy products, either on their own or in combination. Most of the studies discovered in the review were not attempting to answer that question, and limitations to the application of the techniques and/or the data analysis approach and presentation were uncovered. In most instances there was limited discrimination between stimuli using implicit techniques unless the stimuli were very different, and oftentimes the explicit methods were more discriminating. However, it still must be acknowledged that they are not necessarily measuring the same thing. Many studies recommended that combining data from different approaches including both implicit and explicit data is likely to be the best approach for predicting consumer response.

There were limited numbers of emotional measurement studies involving consumed foods with FEA and EDA, and only one for EMG. There is a dearth of papers on dairy and none combing EMG, FEA and EDA. There is therefore both a need and an opportunity to investigate the ability of these techniques either on their own, in combination with themselves, or in combination with explicit cognitive measures of feelings, to discriminate between products and predict affective response. This needs to be done in tandem with the development of improved experimental protocols, and data analysis and presentation techniques. If approaches to measuring emotional response that discriminate between similar products within a category can be found, it would be a valuable step towards developing approaches that can better predict consumer choice.

1.7 Research Objective and Hypotheses

Based on the review of methods discussed above and the time frame of a Masters, this research focused on identifying the benefits and limits of implicit methods of emotion measurement in characterising emotional response (FEA, EMG, EDA) to predict consumer choice and behaviour alongside selected common explicit approaches. The investigation was guided by the following research questions:

- Can milk and yoghurt products be differentiated by selected explicit measures of emotion and liking?
- Can the milk and yoghurt products be differentiated by selected implicit measures of emotion?
- Do the implicit and explicit measures of emotional response and liking correlate?
- Do the implicit and explicit measures of emotional response and liking differentiate milk and yoghurt to different extents?

- Are the selected implicit and explicit measures impacted by the use of an individually composed written evoked scenario?

The key hypotheses were:

1 a: Products within a dairy category can be discriminated by physiological facial EMG measures.

b: Products within a dairy category can be discriminated by physiological EDA measures.

c: Products within a dairy category can be discriminated by FEA measures.

d: Products within a dairy category can be discriminated by RATA cognitive emotion and hedonic measures.

2 a: Implicit measures of emotional response will discriminate products in a consumption context and a lab context, but strength of the discrimination will vary by method.

b: Explicit measures of emotion and hedonic liking response will discriminate products differently when used in a consumption context compared to the lab context.

3: Implicit measures of emotional response will correlate with self-reported measures of emotion and liking but the strength of this relationship will vary across the different measures.

4: Emotional response measured with implicit methods (EMG, EDA, FEA) will discriminate products differently to emotional response measured with an explicit method.

These objectives and hypotheses were considered during experimental design and set up, as detailed in the following section.

2 MATERIALS AND METHODS

This study measured selected implicit and explicit responses of consumers to dairy products using facial electromyography (EMG), electrodermal activity (EDA), facial expression analysis (FEA), and the EsSense 25 emotional lexicon alongside hedonic liking. The effect of evoked scenario was also investigated, with consumers tasting samples in two different sessions, one in the laboratory with no additional scenario and the other where the participant wrote their own relevant scenario.

2.1 SAMPLE INFORMATION

Milk and yoghurt products were preselected to give a range of different flavours and textures within the categories of unflavoured milk and unflavoured natural yoghurt. The milk stimuli were selected to represent a sample set with small differences in sensory characteristics, whilst the yoghurt samples represented larger differences in sensory characteristics. These were chosen to investigate the sensitivity of the measures of emotional response to larger (yoghurt) and smaller (milk) differences in sensory characteristics, a key objective of the study. Ten samples were used in the study, five milks and five yoghurts (Table 1).

Table 1. Milk and yoghurt samples used in the experiment.

	Brand	**Type**
Milk	Anchor	Fresh pasteurised blue top[1]
	Anchor	UHT blue top
	Anchor	Trim milk
	Anchor	Full fat (Silver top)
	Dairy Dale*	Blue top
Yoghurt	Fresh n' Fruity	Greek style[2]
	Fresh n' Fruity	Natural (40% less sugar)
	Gopala	Natural yoghurt
	Puhoi Valley	Authentic Greek natural
	The Collective	Kefir pourable

*From gas station. [1]Dummy sample in the milk set. [2]Dummy sample in the yoghurt set.

The different types of milk included: full fat, skim, UHT whole, an oxidized whole milk (as an "extreme" sample), and a fresh pasteurised whole milk. The oxidised whole milk sample was sourced from a gas station where it had been stored in glass-fronted fridges, leading to light oxidation. The yoghurts included were an authentic Greek, a Greek-style natural, reduced-sugar natural, pourable kefir (as an "extreme" sample) and a natural yoghurt. The fresh pasteurised whole milk and Greek style natural yoghurt were presented an additional time as “dummy samples” being tasted first to remove the effect of the first position (Dorado, Pérez-Hugalde, et al. 2016), and the data was not used in analysis. In total, six milk and six yoghurt samples were evaluated in each session by the participants.

2.2 CONTROLLING SAMPLE TEMPERATURE

Controlling the temperature of the samples during storage and transport was important from a food safety perspective as samples should not be at temperatures above 5°C for more than two hours over the course of experiment. In addition, there is some indication that serving temperature can affect participant emotional response, as found by Singh and Seo (2020) in a study using water served at four different temperatures as a stimuli. Therefore, it was important to investigate how the temperature of the samples changed over storage, transport, and presentation to participants to ensure this was controlled as much as possible.

The laboratory where the experimental sessions were conducted was in a different building to the food grade laboratory where the samples were prepared. Consequently, samples had to be stored in a small fridge in the same building as the data collection laboratory. Before data collection began, the temperature of the fridge where the samples were stored was tracked to measure its variability and ensure that the samples would be kept at an appropriate temperature during storage and would be at a cold temperature (~ 5°C) during tasting.

To determine how the temperature of milk and yoghurt samples would change over time inside the fridge, six milk and six yoghurt samples were placed into 35mL-cups (10g each) and then placed inside a plastic container inside an insulated bag with a frozen ice pack for transport to the data collection laboratory. Upon arrival, the temperature of both sets of samples was measured using a thermocouple and a thermometer, before the thermocouple was placed into the middle of one milk sample through a hole in the centre of the lid. The samples were then placed inside the fridge and the temperature of the milk sample recorded every minute for the next hour. Following this, the samples were removed, and the thermocouple transferred to a yoghurt sample container. When the temperature of the yoghurt reached 6.3°C (same temperature after the transport to the data collection laboratory), the samples were returned to the fridge and the same measurement schedule as for the milk was followed. Over the course of the hour, the temperature of the milk and yoghurt samples decreased from 7.1°C to 1.7°C and 6.3°C to 1.2°C, respectively (Figure 1). Because the temperature of the samples was expected to increase during the time between removal from the fridge and consumption by participants, the temperature setting of the fridge was changed to the coldest for future tests to ensure that the samples would be the coldest possible by the time of removal and presentation to participants.

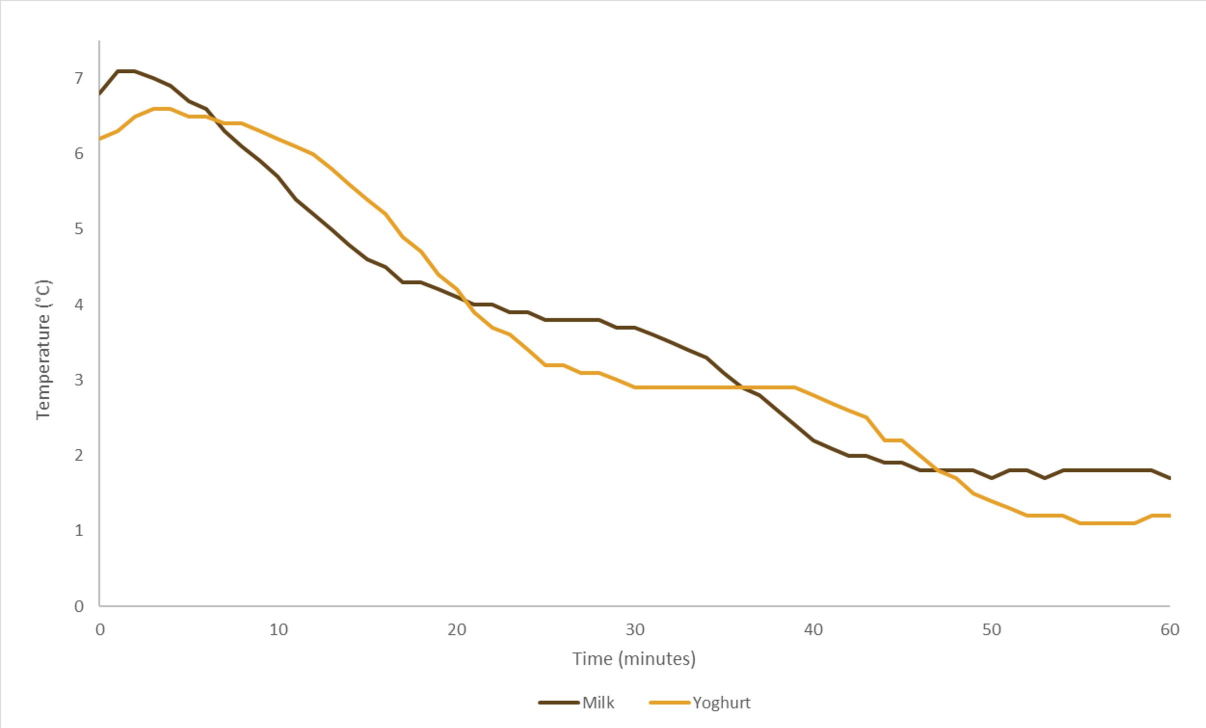

Figure 1: Internal temperature variation of the milk and yoghurt samples inside the fridge over 60 minutes. Time corresponds to after the fridge door was closed.

In order to determine how the temperature of the milk and yoghurt samples would change over the course of an experimental session, a simulated session was run following an estimate of the experimental structure. Six milk and six yoghurt samples (10g each) were placed into 35mL-cups and then placed in the fridge for one hour. Samples were then removed and placed in an insulated bag with a frozen ice pack. After 10 minutes, the milk samples were removed from the insulated bag and the temperature recorded. To understand the amount of time that a participant may spend during the experimental sessions (with and without scenario) a simulation was carried out. The milk samples were left at room temperature for 5 min before the temperature was measured again. This measurement simulated a participant tasting the first milk sample, and the temperature was measured every 2 minutes for the next 10 minutes to represent the time of the participant tasting and answering a questionnaire for the other 5 milk samples. Directly after the last milk sample, the yoghurts were removed from the insulated bag and the temperature recorded using the same protocol used for the milk samples. Figure 2 shows the temperature of the milk and yoghurt samples over 40 minutes (the expected course of the simulated experimental session) increasing from 3.6°C to 11.3°C and 10.8°C, respectively. Thus, it was observed that a container with a better insulation should be used to store the samples during transport. Additionally, to reduce temperature increase, it was decided to present the samples on a plastic tray that had been previously refrigerated and then stored in the insulated container alongside the samples.

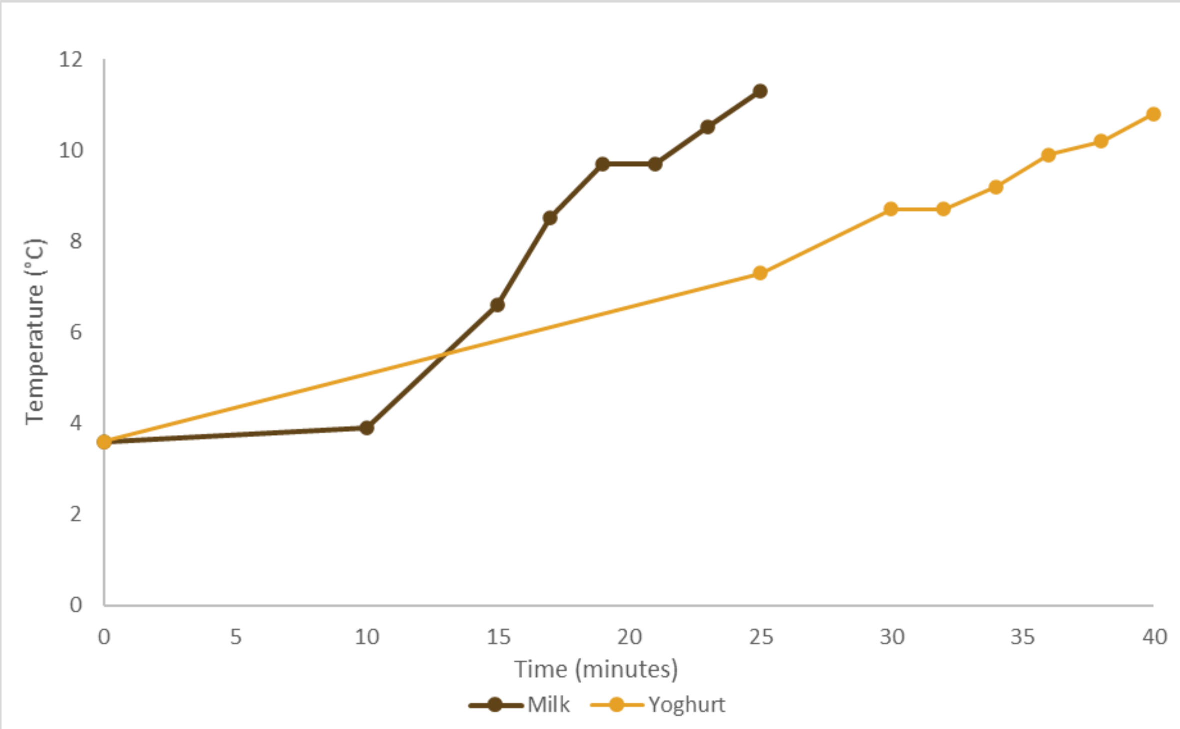

Figure 2: Internal temperature variation of the milk and yoghurt samples inside the fridge over 60 minutes. Time corresponds to after samples were removed from the fridge.

Another pilot session was performed, but this time with the full presentation software in use. The temperatures of a duplicate set of milk and yoghurt samples were measured at the times when the participant would taste the samples. Twelve milk and twelve yoghurt samples were placed into 35mL-cups (10g each) and then placed into two plastic containers, leaving the milk samples stacked on top of the yoghurt samples. The plastic containers were closed and put inside a 25L-insulated container with frozen ice packs for the transport to the data collection laboratory. After one hour in the fridge, the plastic containers were removed and again placed in the insulated container during the 35 minutes of set-up time before presentation. The milk samples were presented first, with temperature of the samples measured when the participant tasted it, which was repeated when the yoghurt samples were presented following the same protocol. At the time of presentation, the milk samples were 3.8°C and by the time the final sample was tasted it measured 10.0°C, whereas the yoghurt samples started at 0.0°C and increased up to 7.6°C (Figure 3). This difference in starting temperature was likely due to the yoghurt's placement on the bottom of the storage container where it was in contact with the frozen ice packs. To reduce the temperature of the milk before presentation, it was decided to place the milk samples on the bottom of the plastic containers during transport and storage, with the yoghurt coming in contact with the ice packs when the milk samples were removed.

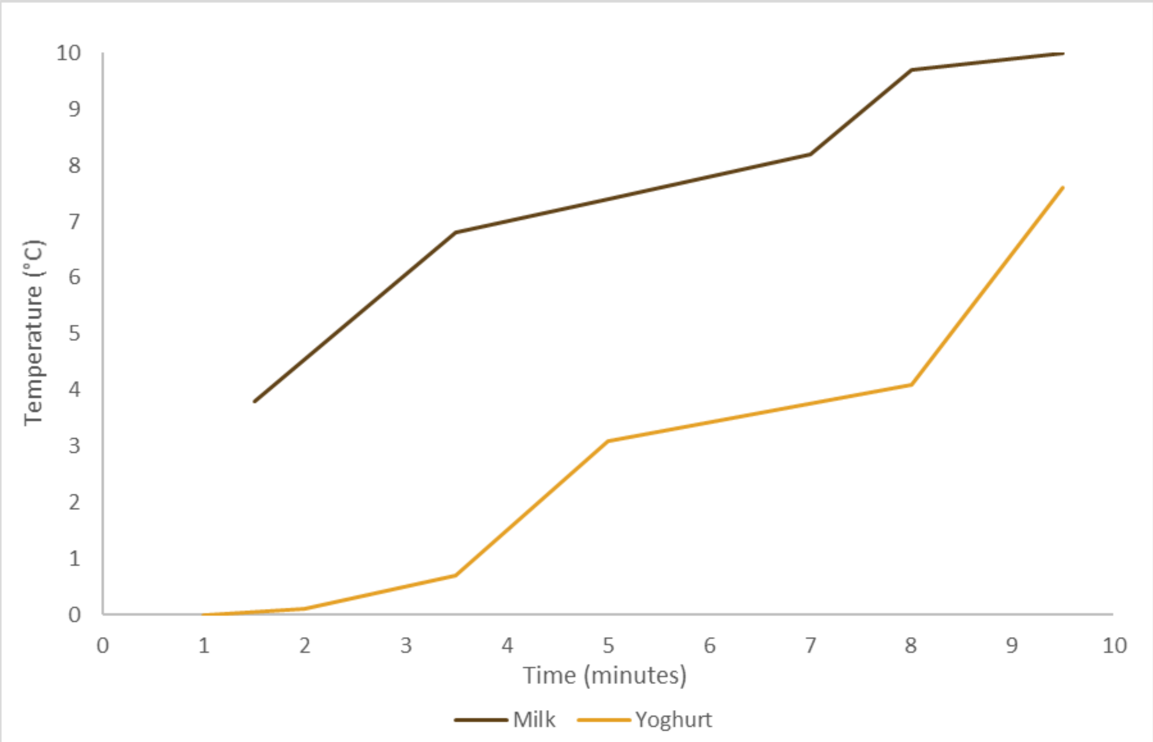

Figure 3: The temperature of milk and yoghurt samples after being presented to a participant.

With the intention to schedule four participant sessions each day, it was important to investigate the effect of storage position of the plastic containers in the insulated container during transport and in the fridge during storage on the temperature of the samples. For this aim, four sets of samples (six milk and six yoghurt samples in each set) were prepared and the temperature recorded before being placed in storage containers. These plastic containers (labelled as A, B, C and D) were placed in the insulated container where containers A and B were directly in contact with the frozen ice packs and containers C and D stacked on top (stage 1, Figure 4). Once at the data collection laboratory, the temperature of the samples was recorded before they were placed in the fridge as shown in stage 2 in Figure 4. After one hour in the fridge, the temperature of the samples was recorded, and then they were placed in the fridge in the other position for an additional 30 minutes (stage 3, Figure 4).

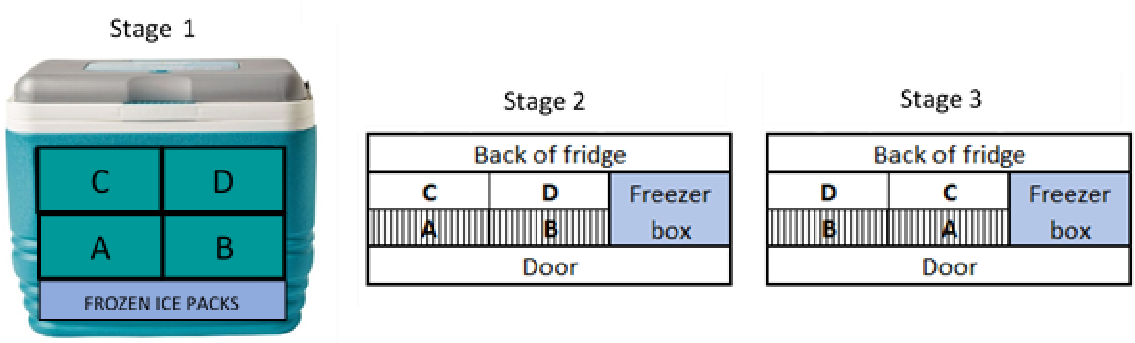

Figure 4: Insulated container layout and positions of the plastic containers (A, B, C, D) inside the fridge.

All samples increased in temperature except the milk samples in container A. This was likely due to the direct contact with the ice packs. The milk samples in container B were also in contact with the ice pack but they had a smaller increase in temperature than those in container C and D. After one hour in the fridge, all milk samples were below 4°C except those in container C (Figure 5), indicating that the samples cooled faster when they were closer to the freezer compartment (see fridge layout in Figure 4). This was supported by the temperature changes in the yoghurt samples, as the containers closest to the freezer compartment (B and D) were below 4°C whereas A and C were above 4°C (Figure 6). Following the second hour of refrigeration, all milk and yoghurt samples were below 4°C, with those closest to the freezer compartment measuring -0.8°C to 0.4°C compared to 2.4°C to 3.0°C in the containers on the opposite side of the fridge. These measurements indicated that the samples placed on the top layer during transport should be positioned on the side of the fridge closer to the freezer compartment to reduce the temperature as fast as possible. It was decided that the first set of samples to be used would be taken from this side, and the remaining containers were to be rotated so that those further away from the freezer compartment were moved closer. Further, the lids of the storage containers were removed before placing them in the fridge in order to increase the air flow around the samples and increase the rate of cooling.

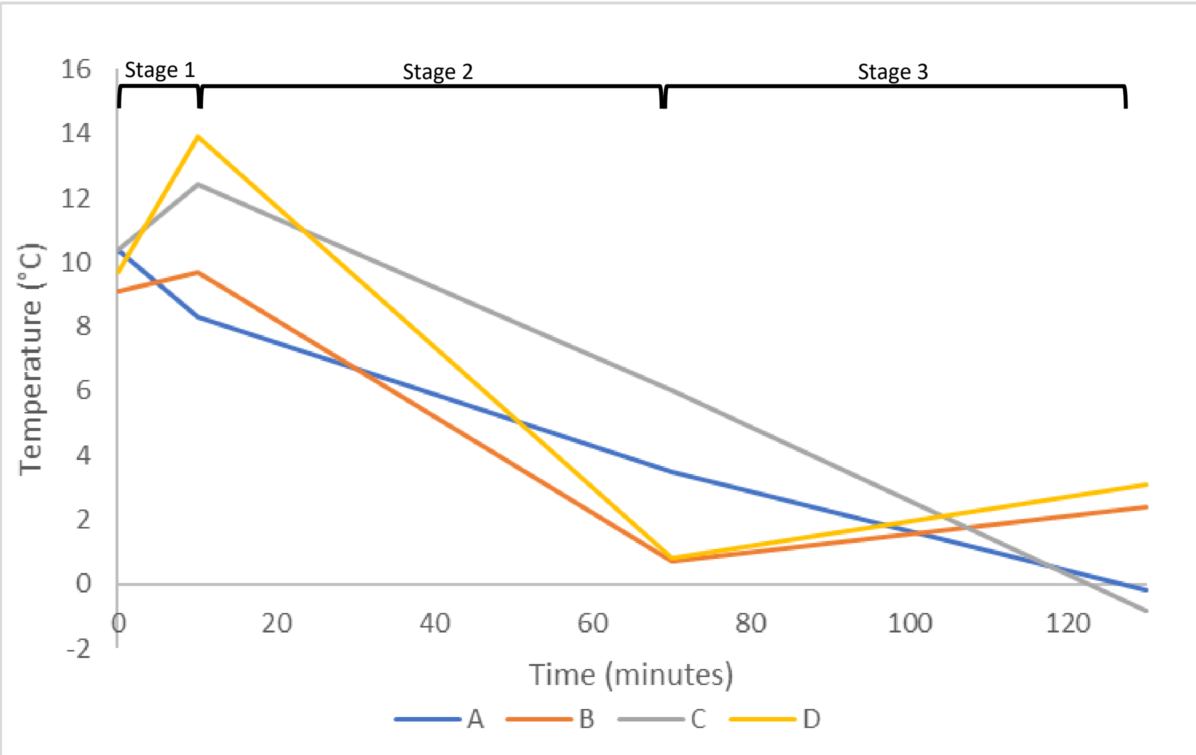

Figure 5: Temperature variation of milk samples placed into four plastic containers (A, B, C, D) over 130 minutes. Time corresponds to after samples were prepared, transport and stored in two different positions in the fridge.

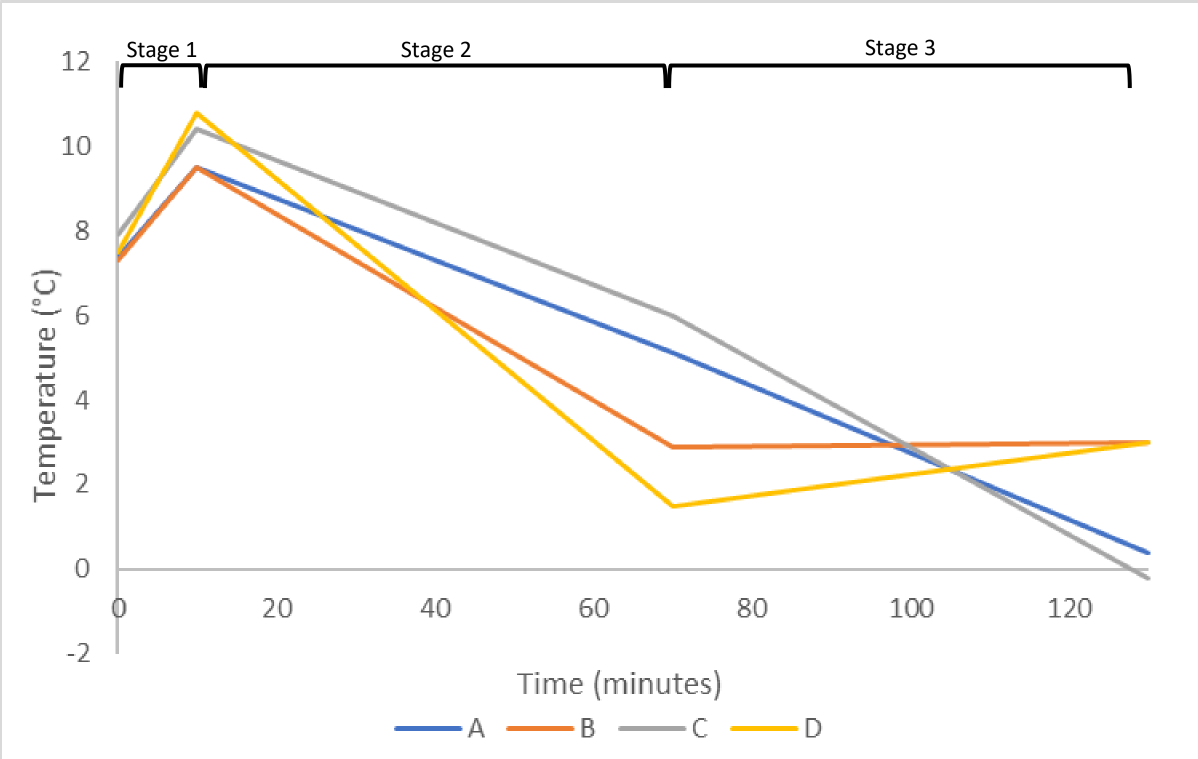

Figure 6: Temperature variation of yoghurt samples placed into four plastic containers (A, B, C, D) over 130 minutes. Time corresponds to after samples were prepared, transport and stored in two different positions in the fridge.

2.2.1 Protocol for transport and storage of samples

From these trials it was determined that the samples would be transported in plastic containers with the milk samples on the bottom, and the yoghurt samples on the top. These containers were to be placed in the insulated container with the frozen ice packs as shown in Figure 4. Before being placed in the fridge, the lids of the containers were removed and the containers that were in contact with the ice packs during transport were placed on the opposite side of the fridge to the freezer compartment. The containers of samples were left in the fridge for at least 1 hour before the first session. Approximately 10 minutes before the session, one container of samples was removed and placed in the insulated container before presentation to participants.

2.3 Understanding Positioning of sensors for EMG and EDA

2.3.1 Facial EMG electrode positioning

Prior any piloting sessions, the application of the facial EMG electrodes was practiced on several volunteers without any data being recorded to develop a clear protocol for the full study. First, the participant was asked to wash their face using a cleanser (Cetaphil oily skin cleanser, Galderma Laboratories, Fort Worth, TX). Two wireless EMG transmitters (BioNomadix 2CH, BIOPAC Systems Inc., Goleta, CA) were then attached to participant's forehead using a 76cm-strap (BioNomadix, BIOPAC

Systems Inc., Goleta, CA). Once the transmitters were secured, the areas for electrode application were cleaned with an alcohol wipe (Medi-Swab, BSN Medical, Luxembourg). Following this step, each of the areas were gently exfoliated using an abrasive pad (ELPAD, BIOPAC Systems Inc., Goleta, CA) to remove any excess dead skin, and subsequently wiped using another alcohol wipe.

The sticker covering the adhesive on one side of two 8mm diameter (ADD208) double-sided adhesive collars (BIOPAC Systems Inc., Goleta, CA) was removed and 8mm EL658 Reusable Ag-AgCl snap electrodes (BIOPAC Systems Inc., Goleta, CA) were attached. This was then repeated with six 4mm diameter (ADD204) double-sided adhesive collars and 4mm EL654 Reusable Ag-AgCl snap electrodes (BIOPAC Systems Inc., Goleta, CA).

Directly prior to application, the electrodes were filled with an electrolyte gel (GEL100, BIOPAC Systems Inc., Goleta, CA) using a 1mL-plastic syringe (Terumo Medical Corporation, Tokyo, Japan), and then a small amount of the electrolyte gel was rubbed into the area where the electrode would be applied to absorb into the skin. The electrode was then placed on the prepared area over the required muscle or muscle group (Figure 7), and a small amount of gel was dabbed onto the back of the sticker to prevent it to sticking to any hair or the wires of the other electrodes.

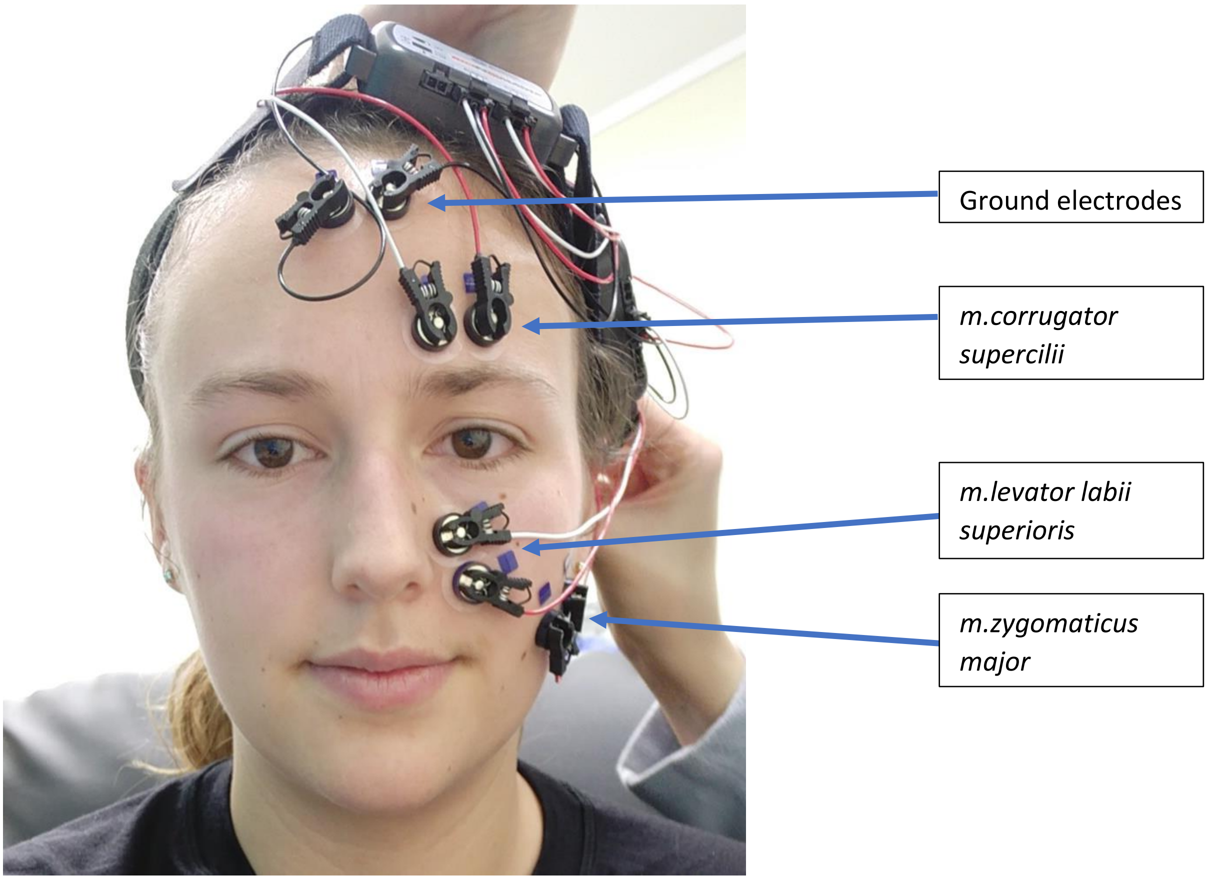

Figure 7: Placement of the EMG electrodes on the face.

Two electrode leads with 3 clips (BN-EL45-LEAD3, BIOPAC Systems Inc., Goleta, CA) were attached to each of the pairs of electrodes positioned over the *corrugator* and *zygomaticus*, with the additional clip on each lead attaching to one of the ground electrodes on the centre of the forehead. An electrode lead with two clips was attached to the electrode pair over the *levator* and impedance was measured using a device (EL-CHECK, BIOPAC Systems Inc., Goleta, CA) to check that it was below 5Ω (signalled as a green light), indicating that the level of noise in the signal was acceptable. If the impedance was greater than 5Ω, the electrode pair was removed, and the skin wiped with another alcohol wipe to remove the gel before the area was gently exfoliated and cleaned with an alcohol wipe. Another pair of electrodes was prepared and filled with gel, before being applied as before and the impedance checked again. Once the impedance was below 5Ω, the leads were plugged in to the transmitters. The *corrugator* and *levator* leads were connected to the same transmitter which used channels 1 and 9 to transmit to the BIOPAC MP160 Data Acquisition System (BIOPAC Systems Inc., Goleta, CA), and the zygomaticus to a separate one which used channel 2.

2.3.2 Electrodermal activity electrode positioning

As with the facial EMG electrodes, the application of the EDA electrodes was practiced on several volunteers before any data was recorded. When these participants were instructed to wash their face, they were also asked to wash their hands with soap and water. After the EMG electrodes were attached, an EDA transmitter (BN-PPGED-T, BIOPAC Systems Inc., Goleta, CA) was attached to one forearm using a 20cm-strap (BioNomadix, BIOPAC Systems Inc., Goleta, CA). The palm-side of two fingers on the same hand were cleaned with an alcohol wipe before application. Once the alcohol had evaporated, two disposable EDA electrodes that were pre-filled with gel (EL507, BIOPAC Systems Inc., Goleta, CA) were applied to the prepared sites. The EDA electrode lead (BN-EDA25-LEAD2, BIOPAC Systems Inc., Goleta, CA) was then attached to the electrodes and taped in place using a micropore surgical tape (3M, St. Paul, MN) to prevent any electrode movement. The wires of the electrode lead were also taped to the back of the hand, and the end of the lead was plugged in to the transmitter on the wrist. The transmitter was connected to the BIOPAC MP160 Data Acquisition System (BIOPAC Systems Inc., Goleta, CA) via Bluetooth on a separate channel to the EMG signals (channel 11).

During the pilot sessions, the placement of the EDA electrodes on the hand and the position on the hand were trialled. It was decided that participants would move their hand that was operating the computer mouse the least, as opposed to that used to lift samples, so this was chosen. The two fingers furthest from the thumb were selected as the location of the electrodes (Figure 8) as this would impede mouse usage the least.

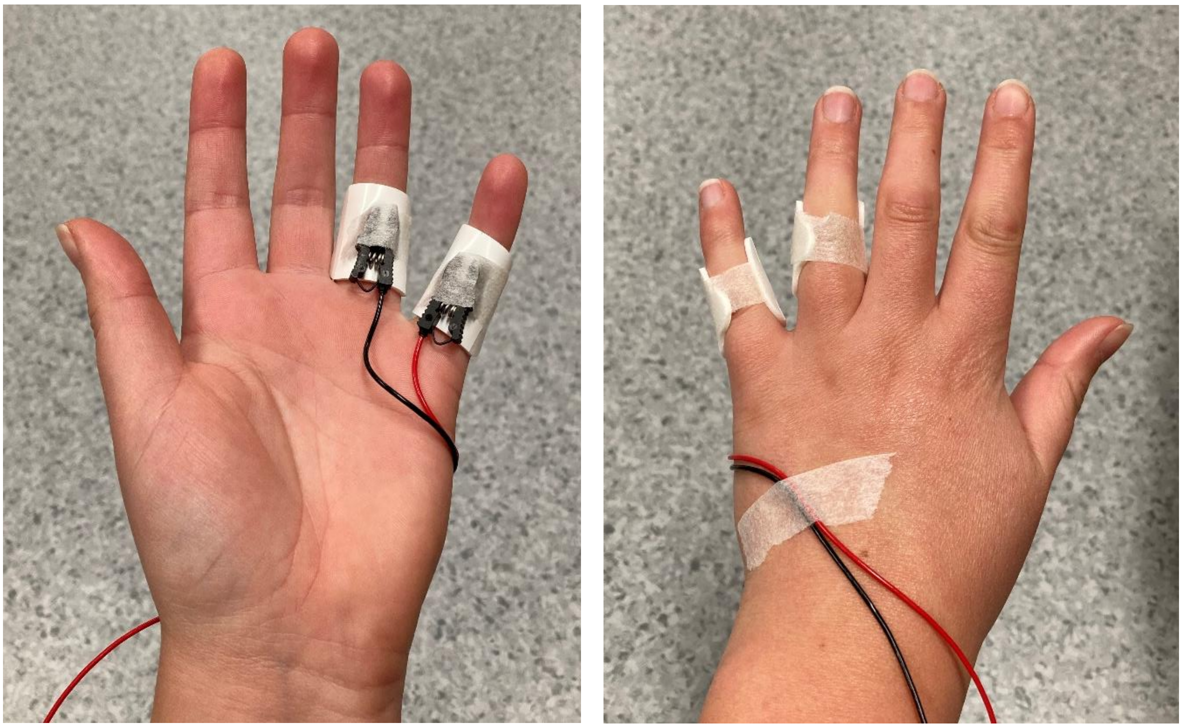

Figure 8: Placement of the EDA electrodes on the participant's hand.

2.3.3 Positioning of the video camera and computer monitor for facial expression analysis

To capture facial expressions, a webcam (C922 PRO HD STREAM WEBCAM, Logitech, California, USA) was placed in front of a 24" computer monitor (Dell™ UltraSharp™ 2407WFP, Dell, Texas, USA) as it needed to be directly facing the participant when they were looking at the screen. As this was obstructing some of the screen from view, the computer monitor was elevated so that the whole screen could be seen by the participant. The computer monitor and camera were positioned approximately 50cm from the edge of the desk, with the angle of the camera adjusted where necessary to centre the participant's face in the frame. The video was recorded through iMotions 8.1 software (iMotions A/S, Copenhagen, Denmark) which would later process the recordings through the AFFECTIVA (AFFECTIVA, Boston, USA) facial expression analysis algorithm to give a measure of facial emotional response for seven emotions (Anger, Contempt, Disgust, Fear, Joy, Sadness and Surprise).

2.4 SET UP OF ROOM

The data collection laboratory had a control and a testing room. The testing room contained the electrode preparation area and the participant desk, and the control room area was where the researcher monitored participant progress through the task (Figure 9). The separate room for the researcher was necessary as the presence of a person sitting next to the participant with no interaction during the task has been shown to affect their facial muscle activity (Nath, Cannon, and Philipp 2019). On the participant desk, there was a wireless keyboard, a mouse, and the computer to navigate through the task with a webcam for recording facial expressions. The fridge was in a different room to avoid heat and noise from the motor disturbing the participants and potential artifacts in the EMG data caused by the electrical motor switching on.

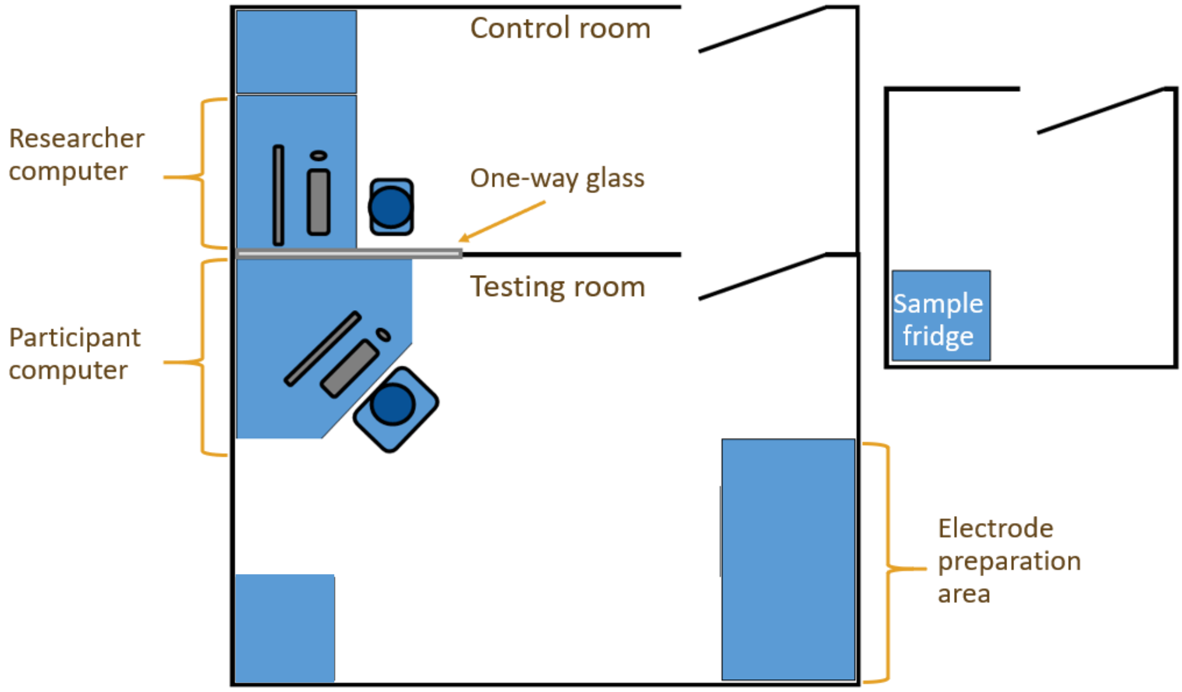

Figure 9: Layout of the data collection laboratory.

2.5 Development of experimental software

A bespoke software presentation application was developed by Malcolm Loudon (School of Psychology, Massey University) using Atom (GitHub, San Francisco, USA), for use on the participant computer during data collection. The software led participants through the experimental session as outlined in Figure 10, and was used to play videos, provide instructions for tasks, and collect participant responses to liking and emotional questionnaires. The sample presentation order for each session was pre-programmed into the software as well as which session each participant would be presented with the need to consider a scenario. The participant's code and session number were entered into the software by the researcher before each session to tell the software which pre-determined order to present the samples (Figure 11).

A second function of the presentation software was to send signals to the computer running the iMotions software when participants reached particular points in the session. These included the start and end of the videos, when the participant entered and left the screens instructing them to taste each sample, and the prompts for making maximum voluntary muscle contractions.

Initially, a shortened version of the software with only one milk and one yoghurt sample was installed onto the computer. This was used to quickly check that the instructions and timings were correct before full piloting. No issues were found with the draft software, so the full version was installed and used in two full pilot sessions detailed in section 2.6 below. During the first of these sessions, the wireless connection between the two computers was disconnected meaning that the signals between the presentation software and iMotions software were not received. For the next pilot and the data collection, the computers were connected using an ethernet cable.

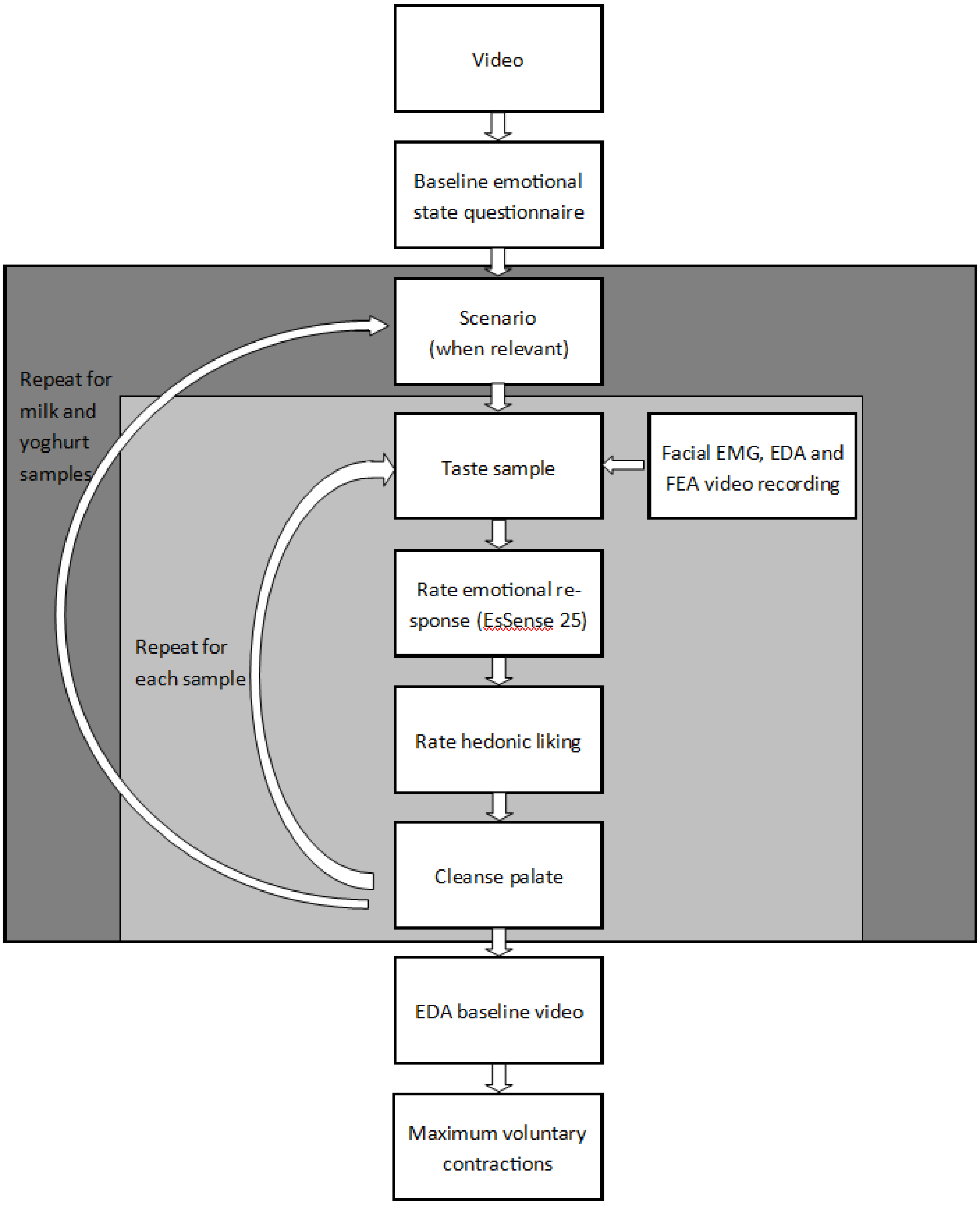

Figure 10: Flowchart of the experimental session showing the stages led by the presentation software.

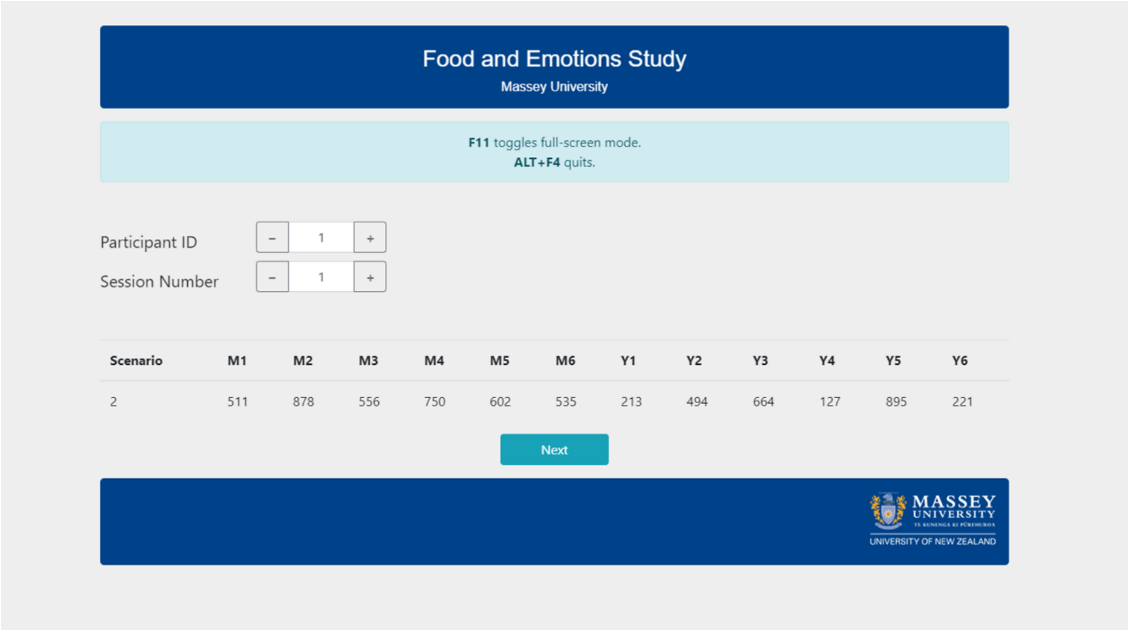

Figure 11: The set-up screen of the presentation software showing the sample presentation order and scenario code for the first session for participant 1.

2.6 PILOT TRIAL OF EXPERIMENT

2.6.1 Samples

Anchor Blue Top milk and Fresh n' Fruity Greek Style yoghurt (Table 1) were used for the pilot sessions. Six 10g samples of the milk were poured into 35mL-plastic cups and labelled with random three-digit codes, and this was repeated with six 10g samples of yoghurt. The samples were transported using an insulated container and frozen ice packs, then stored in the fridge for at least one hour before being removed 10 minutes before the session and stored in the insulated container in the laboratory, according to the protocol developed in section 2.2.1.

Throughout the task, filtered water and plain crackers (Water Crackers Original, Arnott's, Auckland, NZ) were provided as palate cleansers with participants instructed to have a piece of cracker and a sip of water before each sample.

2.6.2 Recording explicit emotional response and liking

Explicit emotional response was measured using the EsSense25 profile (Nestrud et al. 2016) (Table 2). The emotion words were displayed in the presentation software (detailed in section 2.5) as a list where participants could scroll using the mouse wheel and rate the emotion intensity felt from tasting a sample using a 5-point scale (0 = "not at all" to 4 = "extremely"). The emotion words were displayed in a random order which was constant within a session but different for each participant and across each session to prevent any order effects (King and Meiselman 2010). After that, hedonic liking was recorded using an 11-point scale, with the levels displayed from "dislike extremely" to "like extremely" (Figure 12).

Table 2: The EsSense25 emotional lexicon.

Active	Disgusted	Guilty	Mild	Tame
Adventurous	Enthusiastic	Happy	Nostalgic	Understanding
Aggressive	Free	Interested	Pleasant	Warm
Bored	Good	Joyful	Satisfied	Wild
Calm	Good-natured	Loving	Secure	Worried

Figure 12: Hedonic scale used to rate liking after tasting each sample.

2.6.3 Recording implicit emotional response

Implicit emotional response was measured using three methods: facial expression analysis (FEA), facial muscle movement (EMG) and electrodermal activity (EDA). Facial expression analysis was recorded using a webcam and collected through iMotions software. Facial muscle movement and electrodermal activity signals were collected using a BIOPAC MP160 Data Acquisition System (BIOPAC Systems Inc., Goleta, CA) and recorded using iMotions software.

2.6.3.1 Facial electromyography

Facial electromyography (EMG) was used to record muscle activity of the *corrugator*, *levator*, and *zygomaticus* muscles during and after the tasting of each sample. Prior to each session, two 8mm diameter (ADD208) and six 4mm diameter (ADD204) double-sided adhesive collars (BIOPAC Systems Inc., Goleta, CA) were prepared by exposing the adhesive on one side and attaching 8mm EL658 electrodes and 4mm EL654 Reusable Ag-AgCl snap electrodes (BIOPAC Systems Inc., Goleta, CA) to the respective sticker. The cleaning and application steps were followed as described in section 2.3.1 with explanations of each step provided to the participant (see Script 2 in Appendix 3). At the end of the session, the researcher led the participant through a 'maximum voluntary contractions' protocol, as outlined in Script 4 in Appendix 3. This step involved the participants making exaggerated facial expressions of smiling, frowning, nose wrinkling and raising eyebrows, and holding them for 1-3 seconds. The measurements recorded from this task were used to convert the muscle activity signals from the tasting session into a proportion of the maximum voluntary contractions for each specific participant. This allowed for differences in individual muscle structure and movement, as well as any slight differences in electrode placement to be accounted for.

2.6.3.2 Electrodermal activity

The EDA electrode application procedures outlined in section 2.3.2 was followed with explanations given to participants as described in Script 2 of Appendix 3. After tasting the samples, the participants were asked to watch a video (iMotions, personal communication, June 26, 2020) in order to get a baseline measurement of electrodermal activity. This video was intended to give specific valance and arousal responses with images including different scenes: nature, a person walking on a rooftop and a baby laughing.

2.6.3.3 Recording facial expressions

The camera used to record facial expressions was already in position prior to the sessions, as described in section 2.3.3. However, the angle of the camera was adjusted for each session to centre each participant's face in the frame.

2.6.4 Pilot sessions

Two pilot sessions with naive participants were conducted on separate days. During these sessions the scripts in Appendix 3 were followed to ensure that the instructions were complete and made sense. The session began with the participant being greeted in the waiting area and brought to the laboratory before they completed the consent form. The participant was then asked to wash their face and hands before electrodes were applied. Following the application of the electrodes, the researcher left the room and started the recording in the iMotions software. The milk samples were presented once the participant completed the baseline EsSense 25 questionnaire, and the yoghurt samples were presented during the break. Once the last sample had been tasted, the video intended for EDA baseline was played, and then the researcher led the participant through the maximum voluntary contractions procedure. Following this, the recording was stopped, the electrodes removed, participant's face cleaned, and the participants were asked for feedback.

2.6.5 Feedback and alterations

Both of these sessions used the full presentation software, however the wireless connection between the computers was disrupted during the first session and changed to an ethernet connection for the second session as described in section 2.5. In addition, a warning statement was included in the information sheet about 'imagery of heights' after a participant gave feedback that the final video might cause distress to people with a fear of heights.

2.7 Data collection

2.7.1 Participants

The number of participants required for this study was estimated using a power calculation and Monte Carlo simulated data. From this, the minimum number of participants required to show significant differences between similar stimuli was found to be 40. However, to account for the possibility of measures being less discriminating than anticipated and the potential for participant drop-out 80-100 was the initial target for recruitment, but this was reduced to 60-80 due to delays.

Prior to any contact with potential participants, ethics approval was obtained from Massey University Human Ethics Committee Southern A, Application 20/30 (Appendix 1). Volunteers were recruited through the Food Experience and Sensory Testing Lab (Feast) consumer database and by using posters on campus (Figure A1: , Appendix 1). Participants were preselected according to the following criteria: consume milk and unflavoured natural yoghurt at least once a week, not pregnant or lactating, and no allergies to any of the ingredients of the samples (see more details in the Information Sheet in Appendix 2). In addition, potential participants were between the ages of 18 and 65 years to avoid ethical concerns of involving vulnerable populations in consumer testing. Sixty participants took part in the study (21 male and 39 female), with an average age of 38.6 (standard deviation ±10.8 years). Participants were given a 'goody bag' after the first session and a $50 supermarket voucher after the second session as compensation for their time.

2.7.2 Samples

The milk and yoghurt products detailed in Section 2.1 were purchased weekly from local supermarkets. One bottle of each milk was used to prepare the samples throughout the week, with a new bottle opened on the first day of each week of testing. The yoghurt products were less homogenous than the milk, requiring stirring before sample preparation, so a new container of yoghurt was opened each day to minimise differences in sensory properties over different days. All samples were stored under refrigerated conditions (at or below 4°C) from the time of purchase to when presented to the participants.

Samples were stored and prepared in the Product Development (PD) Laboratory in the Riddet Complex at Massey University. For each, 10g of milk or yoghurt were placed in a 35mL-plastic cup labelled with 3-digit code and sealed with a lid. Then, the samples were placed in an insulated container with an ice pack before being placed in the refrigerator near the data collection laboratory in the Psychology building. The samples were then stored there until just before presentation to the participants.

At the start of the experiment, the milk samples (served on a small plastic tray) were placed on the desk in front of the participant, and yoghurt samples were kept in an insulated box with an ice pack to keep it at cold temperature until being served to the participants. Both milk and yoghurt samples were presented at 4°C (±3°C). Filtered water and plain crackers were provided as palate cleansers, as in the pilot sessions, section 2.6.1.

2.7.3 Evoking a scenario

During the experiment participants were asked to provide descriptions of a situation and time of day when they would typically consume milk and unflavoured natural yoghurt. Before the start of the session this part of the task was explained, and an example was given (see Script 3 in Appendix 3). In addition, another example was provided within the on-screen instructions, Figure 13. The participants were given 5 minutes to complete this task and were instructed to keep their own scenario in mind whilst tasting the samples. This approach to evoking a scenario was chosen to reduce the likelihood of having a scenario that was incongruent to some participants, allowing participants to have the opportunity to think about a scenario that is specific and representative to them (Dorado, Chaya, et al. 2016).

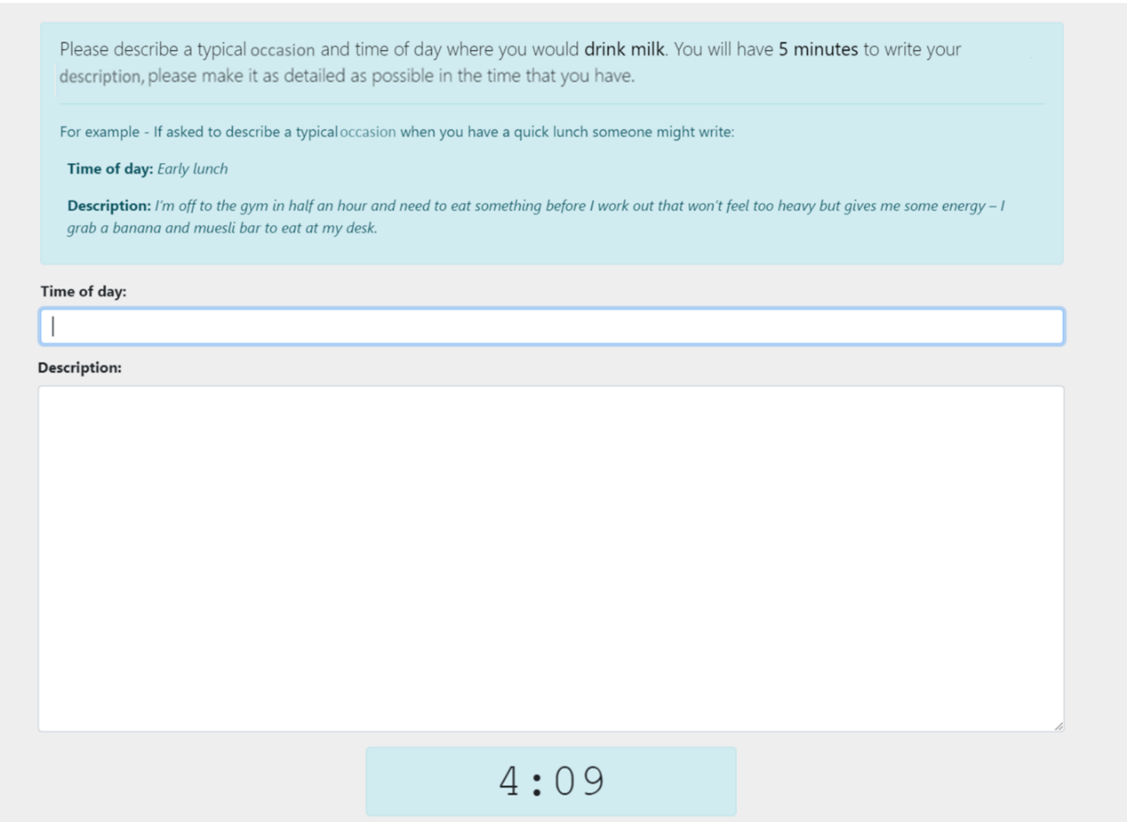

Figure 13: The instructions provided to the participant during the task for recording their evoked scenario for milk.

2.7.4 Recording explicit emotional response and liking

Explicit emotional response and liking were recorded as detailed in the pilot experiment in section 2.6.2.

2.7.5 Recording implicit emotional response

2.7.5.1 Facial muscle activity and electrodermal activity procedure

As in the pilot sessions, facial muscle activity of the c*orrugator*, *levator*, and *zygomaticus* was measured using facial EMG. The preparation and application of the EMG electrodes followed the method used in the pilot sessions (2.6.3.1). The procedure for the application of the electrodes and recording of EDA signals was followed as described in the pilot section (2.6.3.2).

2.7.5.2 Facial expression analysis procedure

The camera location was unchanged from the pilot experiment and the camera angle was adjusted for each participant as described in the pilot (2.6.3.3). During the data collection the lighting in the data collection room was altered as the facial recognition software was having difficulty with the amount of light coming from behind the participant. To reduce the backlighting, the fluorescent ceiling light behind the participant was switched off leaving only the light source above the participant.

2.7.6 Experimental sessions

Each participant attended two sessions which were split across two weeks to ensure that there was at least two days between their sessions. To investigate the effect of the evoked scenario on participant emotional response, only one of the sessions included the scenario element. The first milk sample presented to participants was always the fresh pasteurised whole milk as a dummy sample. The remaining five milk samples (including a second sample of fresh pasteurised whole milk) were served according to a randomised balanced design across participants. Similarly, for the yoghurt samples the participants were presented the Greek-style natural yoghurt first as the dummy sample, and then the next four samples (including the Greek-style natural yoghurt) according to a randomised balanced design. The order of sample presentation was determined using a Williams Latin square design (Williams 1949) to balance first the order of the five milk samples across participants and sessions, and then the order of the four yoghurt samples. However, the kefir sample was always presented last because of its strong flavour that may have influenced the ratings of the other samples that followed it despite the use of palate cleansers.

At the beginning of each session, the participant was met at the waiting area in the Psychology building and then led to the laboratory. Here, the information sheet was further explained, and the consent form signed by the participant before washing their face for electrode application (see Script 1 in

Appendix 3). The EMG and EDA electrodes were applied following the procedures in section 2.7.5.1 above (see also Script 2 in Appendix 3).

The overall structure of the experiment after the application of electrodes is shown in Figure 6. In order to account for differences in emotional state that participants may be in, participants were asked to watch a video clip from Alaska's Wild Denali (Rohlfing 1997) at the start of the task and then answer the EsSense 25 questionnaire to provide a baseline measure of their emotional state. This step was intended to put the participants in a neutral emotional state and to allow the signals from the electrodes time to settle. The participant was then asked to write a scenario or move directly to the sample tasting stage depending on the session treatment (scenario/no scenario). During the sample tasting, participants were instructed to pick up each sample using their hand that did not have electrodes and taste the sample by taking one large sip or large spoonful of it. After tasting each sample, participants were asked to rate their liking and emotional response. Crackers and filtered water were provided as palate cleansers between samples. The EMG, EDA and video data were recorded throughout the task, with signals sent from the presentation software to the iMotions software when the participant was instructed to put the sample in their mouth and when they clicked the button "next" after clearing their mouth. These markers were intended to identify the data needed for data processing, however many participants tasted samples outside of this time frame, so annotations were made in post-processing. After the sample tasting phase, the EDA baseline was measured during a video. Following this, the participants were led through the maximum voluntary contraction procedure by the researcher as detailed in section 2.6.3.1 and Script 4 in Appendix 3.

2.8 DATA PROCESSING

2.8.1 Data processing in iMotions software

After each experimental session, the facial video recordings were run through 'AFFECTIVA post-processing' in iMotions to ensure that all frames of video were processed. After all data collection was complete, annotations were added to the video recording of each session in the iMotions software at the time when the participant started tasting each sample. Specific events such as when the participant touched their face, coughed, or sneezed, and yawned were also added as annotations at this time. Following this, the EDA data was processed in iMotions using the 'GSR Peak Detection' R workbook (iMotions A/S, Copenhagen, Denmark) which finds the peaks in the raw EDA data and separates the phasic (fast changing) and tonic (slow changing) components of the electrodermal activity. All files were then exported from iMotions as CSV files for further processing in R software.

2.8.2 EMG processing in R software

The exported raw EMG signals were processed in R (RStudio, Boston, MA) where they were first filtered through a low-pass (500Hz) filter to remove any frequencies that were out of the normal range of muscle activity (around 100Hz). The data was then passed through a high-pass filter to remove low frequency noise (20Hz) such as movement of the electrodes and wires, and a band-stop filter was run to remove interference from the mains electricity (at 50Hz). A second low pass filter was then run to smooth the data before the signals were rectified which took the absolute value of the signal, converting the negative values to positive. Finally, the maximum voluntary contraction was found for each muscle for each participant, taken from the period at the end of the task where participants were asked to make exaggerated facial expressions. Filtered data for each muscle was then converted to a percentage of the maximum voluntary contraction of that muscle for that participant. This step aimed to reduce the effect of variation in muscle activity between individuals, and variations in electrode placement between sessions.

2.8.3 Sensor processing in R software

For each session, the 10 second period after a sample was tasted by the participant was identified from the annotations made in iMotions and labelled with the sample code. The average of the facial muscle activity (as the percentage of the maximum voluntary contraction), EDA phasic component (measured in microsiemens) and AFFECTIVA emotions (the percentage of trained human coders that would rate an emotion as present) during this period was calculated and combined in a data sheet with the emotion and liking questionnaire responses to each sample for each participant for each session recorded by the presentation software. The milk and yoghurt data sets were saved as different files for ease of analysis.

2.8.4 Issues with facial expression analysis

During the annotation stage of data processing, it was noted that the facial markers that the AFFECTIVA algorithm placed on the video of participant faces to track movements of facial landmarks were often in the wrong position or not recognising a face was present. The markers often focused on the EMG electrodes instead of facial features and also struggled to recognise faces for participants with darker skin tones and when glasses or beards were present. As a result, the processed data had very few values that were greater than zero and the mean values over 10 seconds after the sample was tasted were very small. This indicates that most of the time the software was not identifying any emotions or not identifying that there were faces present. Due to this uncertainty, the resulting data could not be regarded as accurate, so no further analysis was performed on the FEA data. Instead, the application of the technique was evaluated for application to 'tasted' samples to inform future work.

2.8.5 Participant removal

One participant was removed from the data set for all analyses after a review of the video footage showed that they were exaggerating facial expressions during the tasting as the EMG data would not be a true representation.

2.9 DATA ANALYSIS

For each product set, a three-way ANOVA with two-way interactions was performed in SPSS version 27 (IBM, Armonk NY) with product and scenario as fixed variables, and participant as a random factor. Liking, EsSense 25 terms, facial muscles, and EDA were considered as dependent variables. A Tukey post-hoc test was also run alongside the ANOVA (Analysis of Variance) to assess whether the products were different according to each measure (an alpha risk of 0.05 was set as the level of significance for all statistical tests).

Agglomerative hierarchical clustering (AHC) was performed on the data for the liking, EsSense 25 terms, activity of the facial muscles and EDA using XLStat (Addinsoft, Paris). For each measure, the mean scores for each participant across both sessions were the observations, and Euclidean distance and Ward's agglomerative method were used to cluster the participants based on the observations. This enabled variation across participant to be better understood.

Correlation analysis between the different measures (excluding facial expression analysis) was performed in R. First, the z score for each measure was calculated followed by the Mahalanobis distance to identify and remove any outliers. The Pearson correlation between each measure was then calculated and presented in a correlation plot.

Fixed effect plots were created in R to show the ability of the implicit measures to predict the explicit measures of emotion and liking. This used linear mixed models to compute the point estimates and 95% confidence intervals for the EsSense 25 terms and hedonic liking predicted by each facial muscle independent from the others. This was necessary as facial expressions are made up of the movement of multiple different muscles, and muscles can be used in the formation of many facial expressions (Ekman, Friesen, and Hager 1978).

3 RESULTS

3.1 ASSESSMENT OF THE ABILITY OF IMPLICIT AND EXPLICIT MEASURES OF EMOTION AND LIKING TO DIFFERENTIATE MILK PRODUCTS

The results of three-way ANOVA for each measure of emotion and liking for milk is shown in Table 3 below. There are many statistically significant main effects (particularly of participant), these will be discussed in the sections that follow.

Table 3: p-values for the main effects from the three-way ANOVA for each of the measures of emotion and liking for milk with significant effects in bold.

Measure	Product	Scenario	Participant	Product* Participant	Scenario* Participant	Product* Scenario
Hedonic liking	**0.001**	0.434	**0.000**	**0.000**	0.096	0.659
Active	**0.041**	0.723	**0.000**	**0.000**	**0.000**	0.436
Adventurous	**0.024**	0.092	**0.000**	**0.000**	**0.000**	0.223
Aggressive	0.112	0.256	**0.003**	**0.003**	**0.024**	**0.000**
Bored	**0.001**	1.000	**0.000**	**0.001**	**0.000**	0.868
Calm	**0.004**	0.803	**0.000**	**0.003**	**0.000**	0.791
Disgusted	**0.020**	0.416	0.148	0.000	0.000	0.202
Enthusiastic	0.091	0.799	**0.000**	**0.000**	**0.000**	**0.014**
Free	**0.008**	0.810	**0.000**	**0.000**	**0.000**	0.835
Good	**0.006**	**0.047**	**0.000**	**0.000**	**0.000**	0.530
Good natured	0.101	0.229	**0.000**	**0.001**	**0.000**	0.266
Guilty	0.204	0.493	**0.000**	0.831	0.937	0.797
Happy	**0.002**	0.087	**0.000**	**0.000**	**0.000**	0.907
Interested	**0.011**	0.259	**0.000**	**0.001**	**0.000**	0.463
Joyful	**0.030**	0.194	**0.000**	**0.001**	**0.000**	0.267
Loving	**0.016**	0.149	**0.000**	**0.001**	**0.000**	0.759
Mild	0.488	0.494	**0.000**	0.077	**0.000**	0.154
Nostalgic	0.252	0.953	**0.000**	**0.009**	**0.000**	**0.014**
Pleasant	**0.024**	0.359	**0.000**	**0.000**	**0.000**	0.218
Satisfied	**0.001**	0.635	**0.000**	**0.000**	**0.000**	0.905
Secure	**0.009**	0.166	**0.000**	**0.010**	**0.000**	0.864
Tame	0.115	0.423	**0.000**	**0.003**	**0.000**	0.543
Understanding	0.256	0.499	**0.000**	0.140	**0.000**	0.640
Warm	0.067	0.436	**0.000**	0.065	**0.000**	0.204
Wild	0.104	0.743	**0.000**	**0.000**	**0.000**	0.493
Worried	0.351	0.859	**0.009**	**0.046**	**0.000**	0.362
corrugator	**0.002**	0.704	**0.000**	**0.023**	**0.000**	0.843
zygomaticus	**0.034**	0.114	**0.000**	0.432	**0.000**	0.832
levator	**0.001**	0.157	**0.000**	**0.000**	**0.000**	**0.032**
Phasic EDA	**0.050**	0.222	**0.000**	0.570	0.990	0.844

3.1.1 Hedonic liking of milk products

There was a significant effect of product on the liking of milk as well as a significant effect of participant and product*participant interaction ($p<0.05$). Tukey post-hoc test identified two subsets (where overall the Gas Station Blue Top Milk and the Anchor Trim Milk scored lower than the other products), however these subsets masked a more interesting interpretation due to participant*product interaction. Most notably, three clusters of consumers were identified representing different liking patterns for milk products, linked to differences in liking of the Anchor Trim Milk, Anchor UHT Blue Top Milk and Gas Station Blue Top Milk products (Figure 14). Cluster 1 (n=23) was characterised by the highest liking ratings for the Anchor UHT Blue Top Milk and a dislike of Trim Milk; cluster 2 showed lower liking ratings in general but rated Blue and Silver top higher than UHT Blue and Gas Station Blue (n=22) which they neither liked nor disliked, they too dislike the trim milk most; cluster 3 (n=14) was characterised by a clear dislike for the Gas Station Blue Top Milk and in contrast to the other clusters, a liking of Anchor Trim Milk.

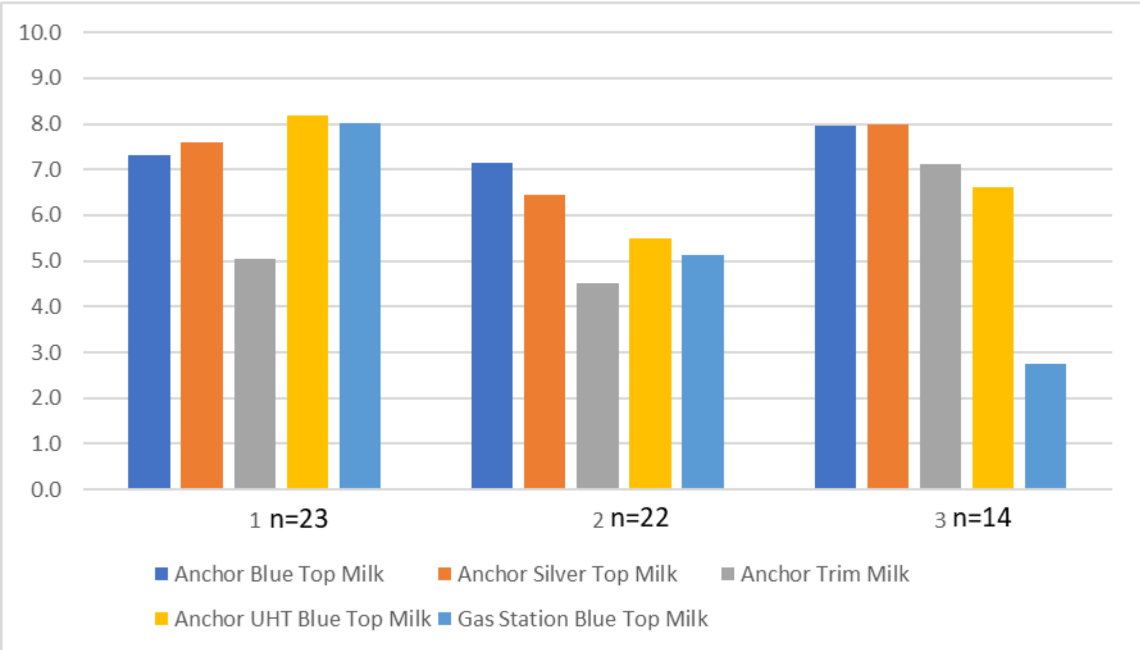

Figure 14: Mean rating of hedonic liking (on a 0-10 scale) for each milk product for the three clusters of participants grouped by their ratings of hedonic liking for milk.

3.1.2 Ratings of EsSense 25 lexicon terms for milk products

Overall, the emotional engagement of participants with ‘milk’ was low, with 19 emotions (Active, Adventurous, Aggressive, Bored, Disgusted, Enthusiastic, Free, Good natured, Guilty, Joyful, Loving, Mild, Nostalgic, Secure, Tame, Understanding, Warm, Wild, Worried) receiving mean scores between 0 and 1, and six (Calm, Good, Happy, Interested, Pleasant, Satisfied) between 1 and 2. These terms

(Calm, Good, Happy, Interested, Pleasant, Satisfied), and another eight (Active, Adventurous, Bored, Disgusted, Free, Joyful, Loving, Secure) had a significant product effect (<0.05), Table 3.

Five emotion terms discriminated the products differently to liking, with only two of these considered both statistically significant and relevant to most participants (Good and Interested). On average, participants rated Good, Interested, and Joyful lowest for the Anchor Trim Milk and rated the Gas Station Blue Top Milk in a similar way to the products that were liked more (such as Anchor Blue Top Milk). Participants felt more Bored tasting the Anchor Trim Milk than the other products, grouping the Gas Station Blue Top Milk with the more liked products. Similarly, participant ratings of Disgusted could differentiate the Gas Station Blue Top Milk from the other products but not the Anchor Trim Milk which was rated similarly for liking.

Closer inspection revealed that not all these terms were used by many participants, in fact every emotion term except for Disgusted had a significant participant, and consequently a cluster analysis was run on all terms. This analysis showed that, excepting Satisfied, there was a cluster of consumers who generally scored emotions lower than “1” (“slightly”) for all products, with cluster size varying by emotion (n=23 to 56), (Figure A2 to Figure A23 in Appendix 4). Terms were considered ‘relevant to most participants’ when more than 50% of participants rated the term above “1” for at least one product and considered ‘significant’ when there was a significant effect of product ($p<0.05$) from the ANOVA. According to such criteria, nine terms were both statistically discriminating and relevant for milk products to most participants: Active, Calm, Good, Happy, Interested, Loving, Pleasant, Satisfied and Secure (Figure 15). A further five terms had a significant product effect ($p<0.05$) but were not relevant to many participants: Adventurous, Bored, Disgusted (Figure 16), Free and Joyful.

Table 4: Groups of EsSense 25 terms based on significant main effect of product ($p<0.05$) and relevance for milk. Terms were considered relevant when at least 50% of participants were in clusters that had mean scores > 1 (“slightly”) for at least 1 product.

	Not significant	**Significant**
Not relevant to most participants	Aggressive, Good natured, Guilty, Mild, Nostalgic, Tame, Understanding, Warm, Wild, Worried	Adventurous, Bored, Disgusted, Free, Joyful
Relevant to most participants		Active, Calm, Enthusiastic, Good, Happy, Interested, Loving, Pleasant, Satisfied, Secure

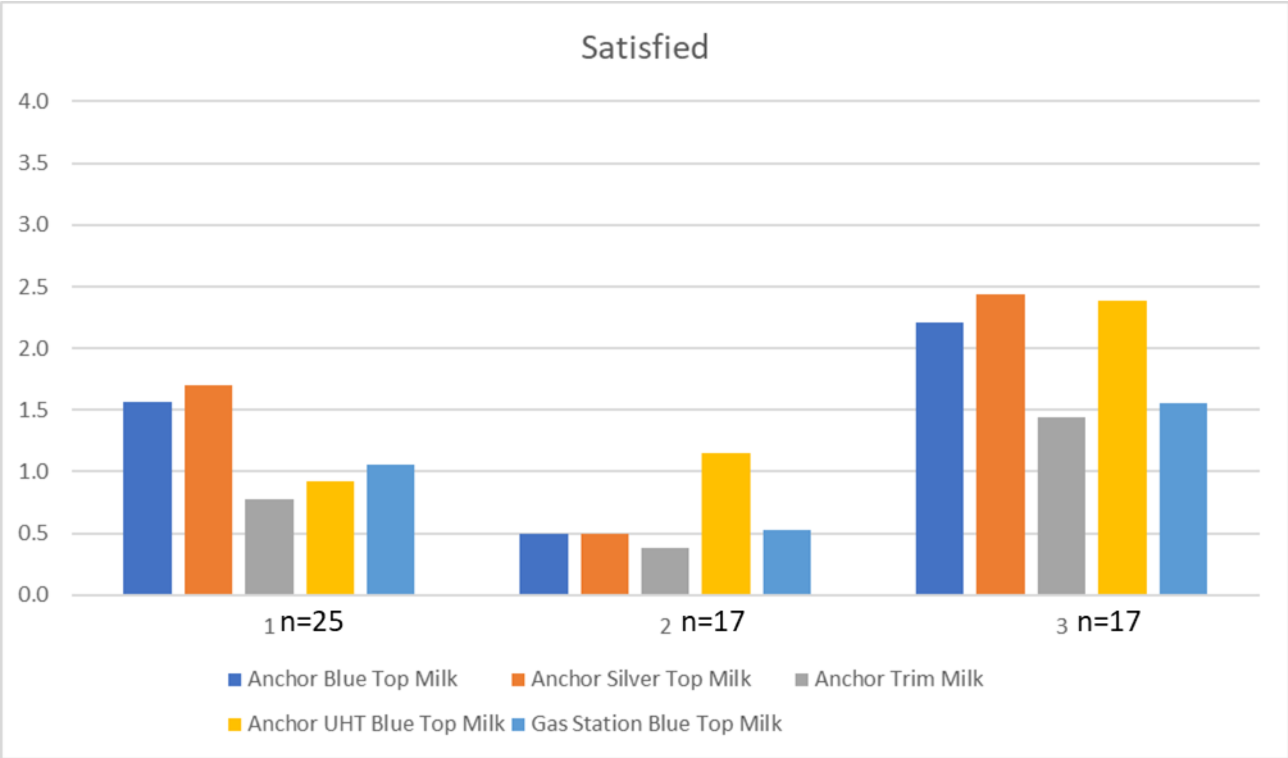

Figure 15: Mean rating of Satisfied (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Satisfied' for milk.

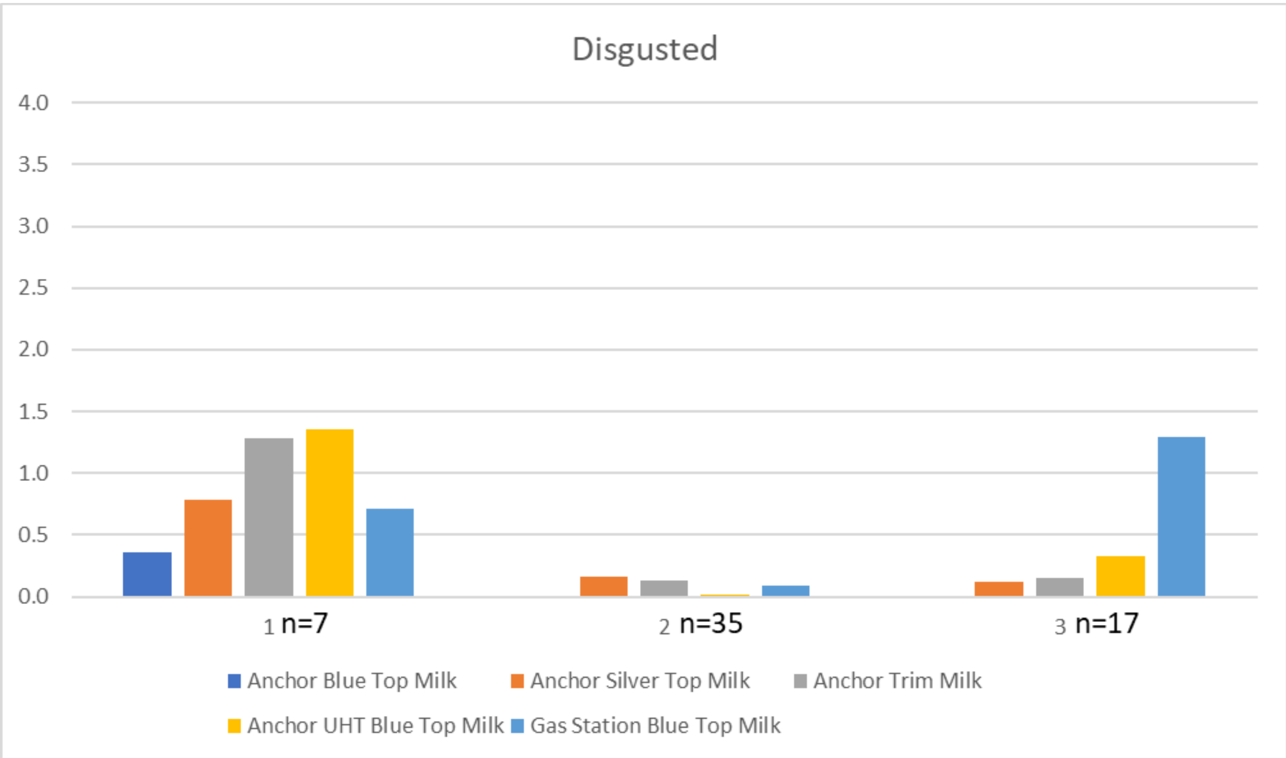

Figure 16: Mean rating of Disgusted (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Disgusted' for milk.

The only EsSense 25 term to be significantly affected by scenario was Good. There was a cluster of consumers (cluster 1: n= 27) who rated Good higher for four of the five products with the scenario than without. There was another cluster (cluster 3: n=9) where the participants rated the Anchor Silver Top Milk higher for 'Good' without the scenario (Figure 17).

Figure 17: Mean rating of Good (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Good' for milk samples across the sessions with (S) and without (NS) the scenario.

3.1.3 Facial muscle movements in response to consuming milk products

For the *corrugator*, participant 18 was removed from the analysis as there was a 17% difference in the mean *corrugator* activity between sessions with and without the scenario indicating that this was an outlier. For the remaining 58 participants, significant product and participant effects were revealed for *corrugator* activity and significant product*participant and scenario*participant ($p<0.05$) interactions. Tukey post-hoc test identified two subsets (where overall the Gas Station Blue Top Milk had greater activity than the Anchor Blue Top and Anchor UHT Blue Top), however these subsets did not account for participant*product interaction. Cluster analysis of participant corrugator activity revealed four clusters (Figure 18), with the largest cluster (n=43) showing little difference in mean corrugator activity across the products. The remaining three clusters were small with cluster 2 representing 4 people with increased muscle activity overall and cluster 3 medium level. Both these two clusters saw more activity for the Anchor Trim and Gas Station milks. Cluster 4 was only two people but saw a different pattern of response with more activity for the Silver top and Gas Station milk.

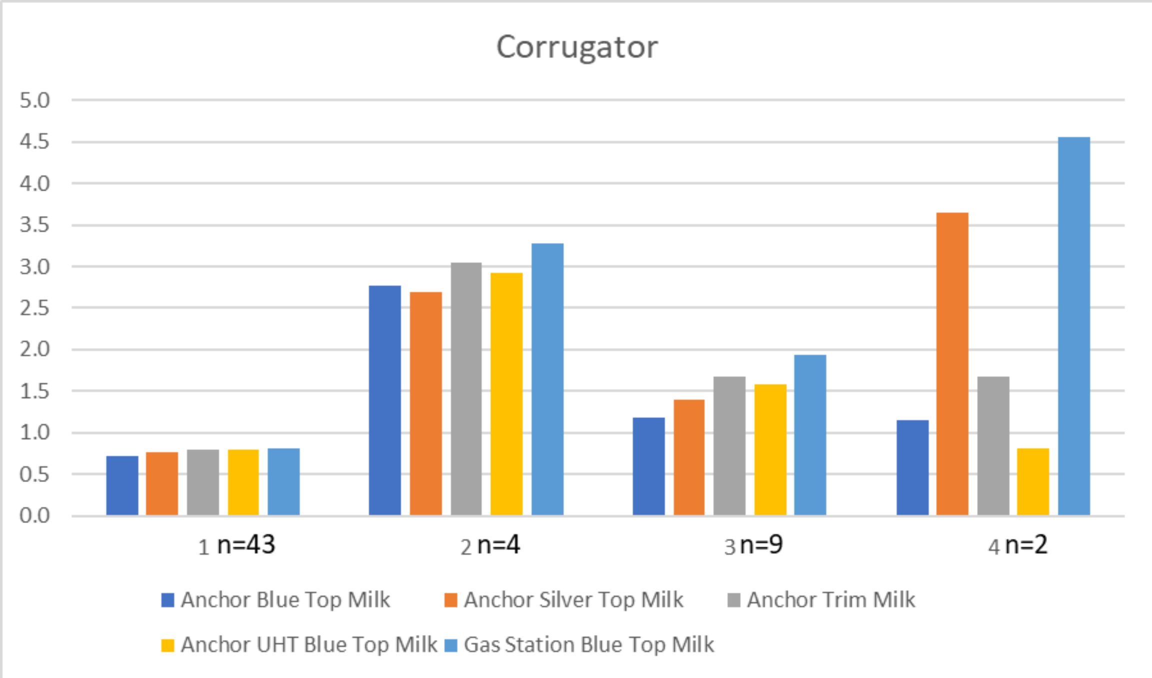

Figure 18: Mean corrugator activity (percentage of the maximum) for each milk product for the four clusters of participants grouped by their corrugator activity for milk.

There were no outliers in the *zygomaticus* data, and whilst there was a significant product effect the Tukey test did not discriminate (all products were in the same set), meaning that the product main effect was likely caused by the participant and/or participant*product interaction ($p<0.05$). A cluster analysis revealed four clusters (Figure 19), with the majority of participants (cluster 1: n=35) showing a mean *zygomaticus* activity of only 1% of their maximum for all products. Cluster 2 showed increased muscle activity to all products but similar lack of product discrimination. Cluster 3, only 5 people, were differentiated by higher muscle activity but also differentiation of UHT and Gas Station Blue. Cluster 4 showed a similar pattern to Cluster 3 but with lower muscle activity in general.

Participant 21 was removed from the *levator* dataset due to a 17% difference in mean *levator* activity between sessions indicating that this was an outlier. For the remaining 58 participants, there was a significant main effect of product and participant on *levator* activity alongside significant participant*product, scenario*participant, and product*scenario interactions ($p<0.05$). Tukey post-hoc test identified two subsets with the Gas Station Blue Top having higher *levator* activity than the remaining products. In addition, cluster analysis was conducted because of the significant participant effect and participant*product interaction, revealing three clusters, (Figure 20). Cluster 1 (n=21) showed little difference between the products with muscle activity of 0.6% of the maximum across all products. The 21 participants in cluster 2 had higher *levator* activity (ranging 1.3% to 1.5%) and differentiated Gas Station Blue. Cluster 3 (n=16) again had the highest muscle activity (2.1% to 2.7%) and differentiated Anchor Blue and Gas Station Blue.

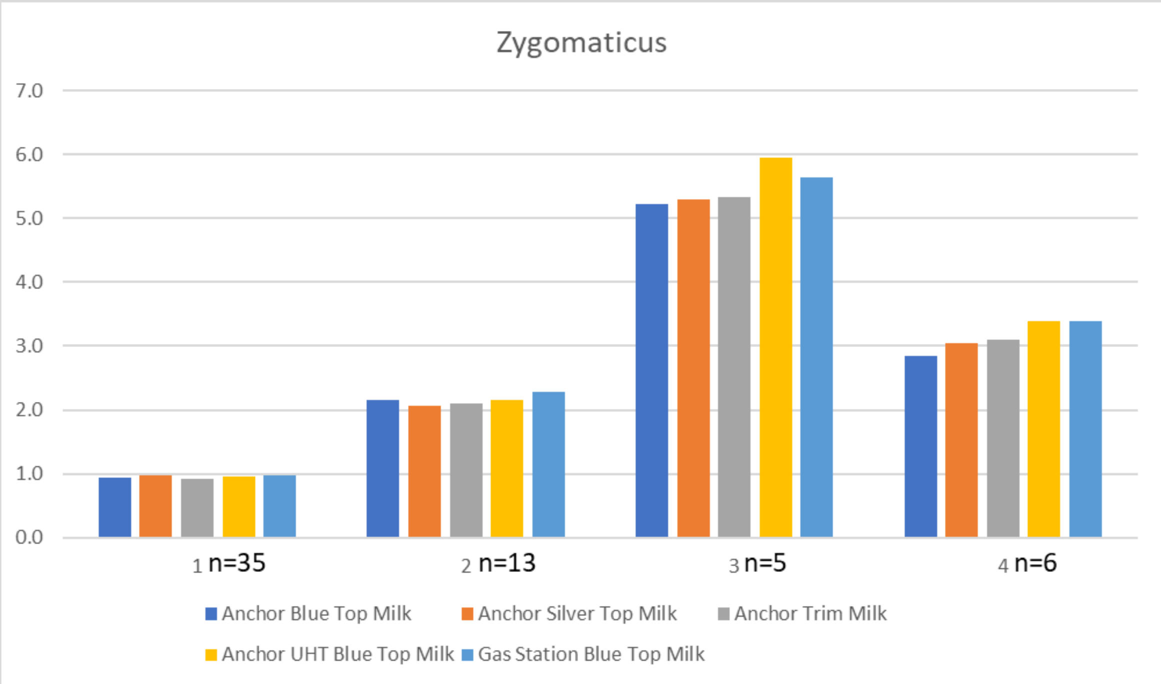

Figure 19: Mean zygomaticus activity (percentage of the maximum) for each milk product for the four clusters of participants grouped by their zygomaticus activity for milk.

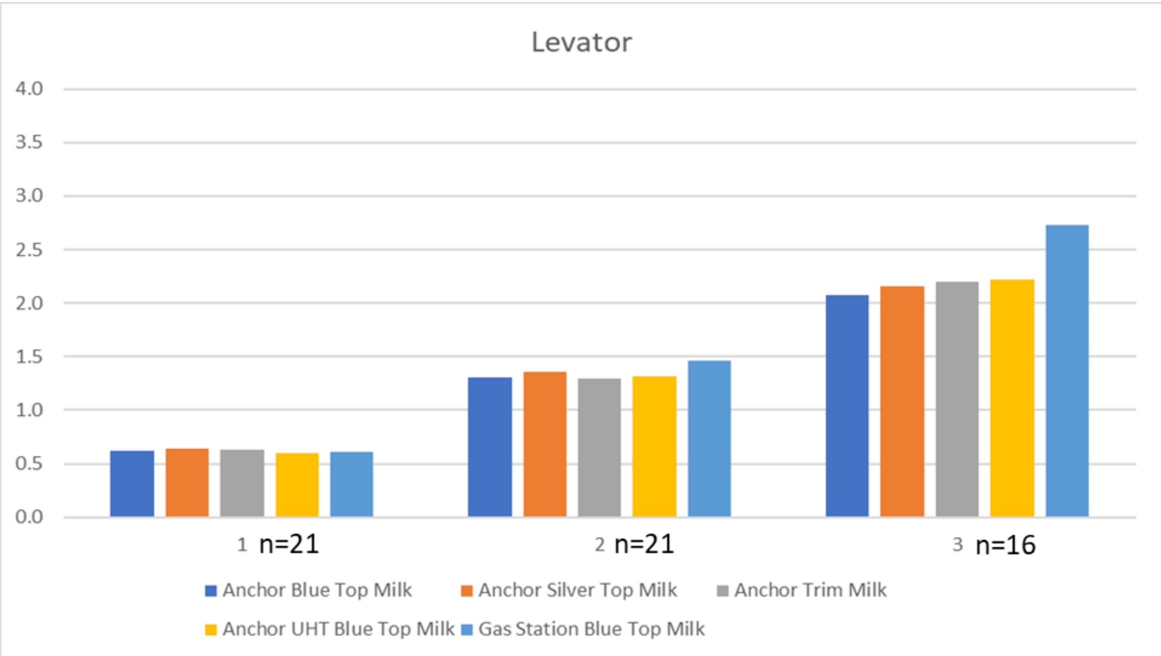

Figure 20: Mean levator activity (percentage of the maximum) for each milk product for the three clusters of participants grouped by their levator activity for milk.

3.1.4 Phasic EDA response to consuming milk products

Product and participant had significant effects on phasic EDA ($p<0.05$), with no interactions between main effects. Cluster analysis gave four clusters (Figure 21), with the two largest showing less than 0.002 micro siemens across all products (cluster 1: n=21 and cluster 2: n=31). The product effect was driven by seven participants across clusters 3 and 4. Cluster 3 (n=6) had the highest EDA for the Gas Station Blue Top Milk (around 0.008 micro siemens), and a slightly higher activity for the Anchor Trim Milk than the other products. For cluster 4 (n=1), the Anchor Blue Top Milk had the greatest EDA (around 0.016 micro siemens) and the activity varied across the other products.

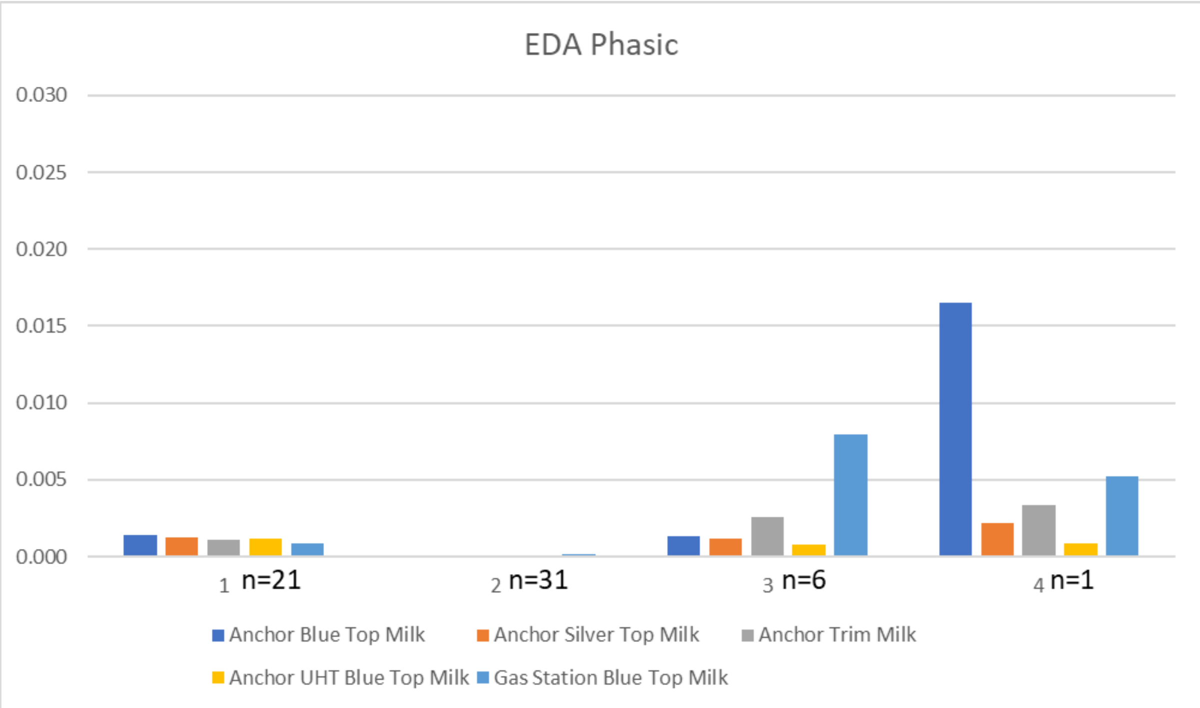

Figure 21: Mean electrodermal activity (in micro siemens) for each milk product for the four clusters of participants grouped by their electrodermal activity for milk.

3.1.5 Correlations and predictive ability of implicit and explicit measures of emotion and liking for milk products

There were no direct correlations between the implicit measures (EDA and EMG) and explicit measures of emotional response (EsSense25) and liking, seen in Figure 22 as correlation coefficients between -0.2 and 0.2. However, there were correlations between hedonic liking and specific EsSense 25 terms: Good, Happy, Pleasant, and Satisfied which all had correlation coefficients of 0.6. These lexicon terms correlated with each other, with coefficients ranging from 0.6 to 0.8. Other interesting correlations between the lexicon terms include a positive correlation between Aggressive and Wild, and a positive correlation between Disgusted and Worried, both with correlation coefficients of 0.6. It should be noted, however, that some terms were rarely used by participants, so some correlations may be due to a small number of participants using the terms rather than the whole population.

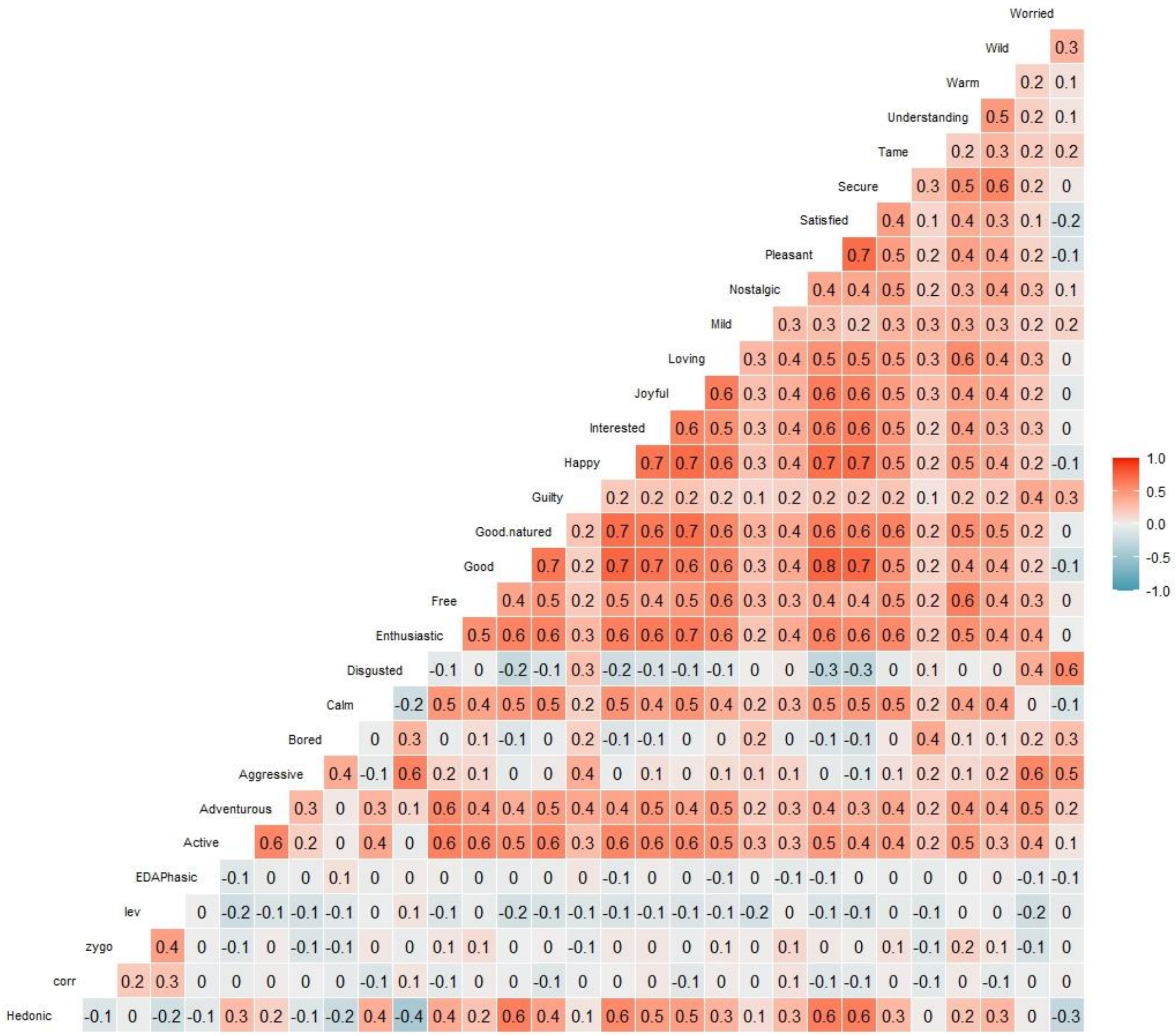

Figure 22: Pearson correlation coefficients for the hedonic liking, implicit and explicit emotion measures for milk samples.

The fixed-effect plots in Figure 23 to Figure 30 show the predictive ability of the EDA and EMG for EsSense 25 terms and hedonic liking. Unlike the correlation plot above, the fixed-effect plots for the EMG look at the predictive ability of each muscle in isolation by removing the effect of the other two muscles. The plots show the change in the score for each lexicon term (or hedonic liking) when a 1-point increase in muscle activity or EDA occurs, with significant effects being ones where the 95% confidence interval does not overlap with the centre line. For example, in the fixed-effect plot of the corrugator and EsSense 25 profile (Figure 23), a 1-point increase in *corrugator* activity (1% of the maximum) predicts a 0.2-point decrease in ratings of Good, meaning that a 5% increase in *corrugator* activity would see a 1-point decrease. The *corrugator* predicted the most terms, having a positive relationship with Aggressive, Disgusted and Worried, and a negative relationship with 11 terms (Calm, Enthusiastic, Good, Good natured, Happy, Interested, Loving, Pleasant, Satisfied, Secure, Warm). The strongest relationships were with Calm, Good, Satisfied and Disgusted, which experienced a change in rating of at least 0.2 points when the *corrugator* activity changed by 1%. The *levator* predicted eight terms (Figure 25), having a positive relationship with Disgusted and a negative relationship with the remaining seven terms (Enthusiastic, Good, Happy, Interested, Pleasant, Satisfied, Warm). The largest effects of *levator* activity on ratings were with Good and Pleasant, where a 1% increase in *levator* activity predicted an approximately 0.2-point decrease in rating. Of the muscles measured with EMG, the *zygomaticus* predicted the fewest terms, having a positive relationship with four terms: Enthusiastic, Pleasant, Understanding and Warm. The predictive ability of this muscle was less than the other two, as the largest effect of a 1-point change in muscle activity was less than a 0.1-point change in rating.

Liking could not be predicted by the *zygomaticus*, however, both the *corrugator* and the *levator* were significant predictors of hedonic liking, with a 1% increase in muscle activity causing a decrease in the liking ratings by 0.5 and 0.7 points, respectively.

Whilst Figure 26 appears to show that EDA can predict changes in ratings of Active, Happy and Pleasant, this does not consider that the EDA values throughout the task are very small (Figure 21) and therefore, a 1-point increase in EDA activity is much larger than what could be expected. In fact, a 1-point increase in EDA would predict an approximately 30-point decrease in the rating of Active, which is not possible on a 5-point scale. Therefore, EDA was not considered to be a significant predictor of any terms.

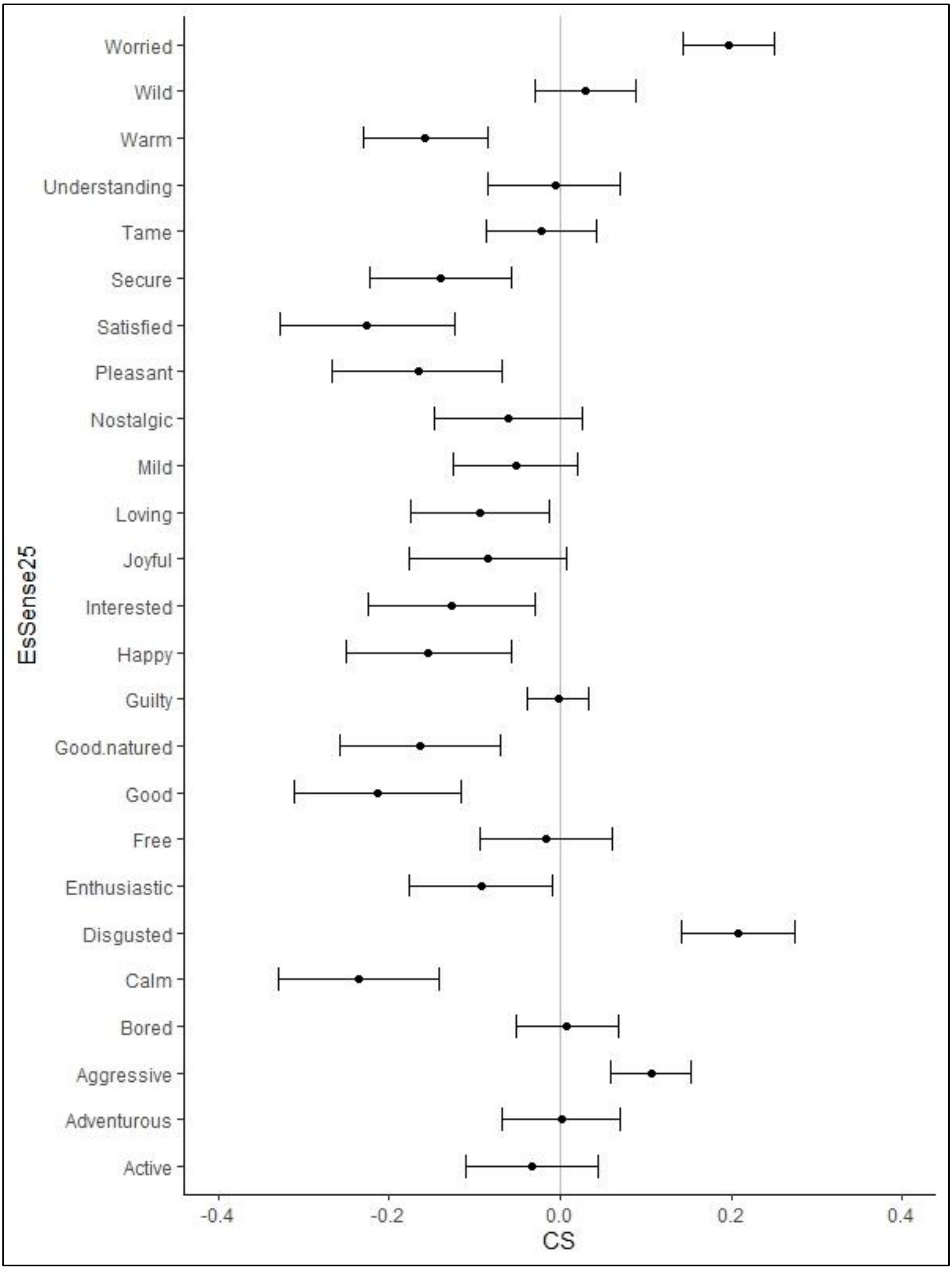

Figure 23: Fixed-effect plot showing the predictive ability of corrugator activity for the EsSense 25 profile for milk samples.

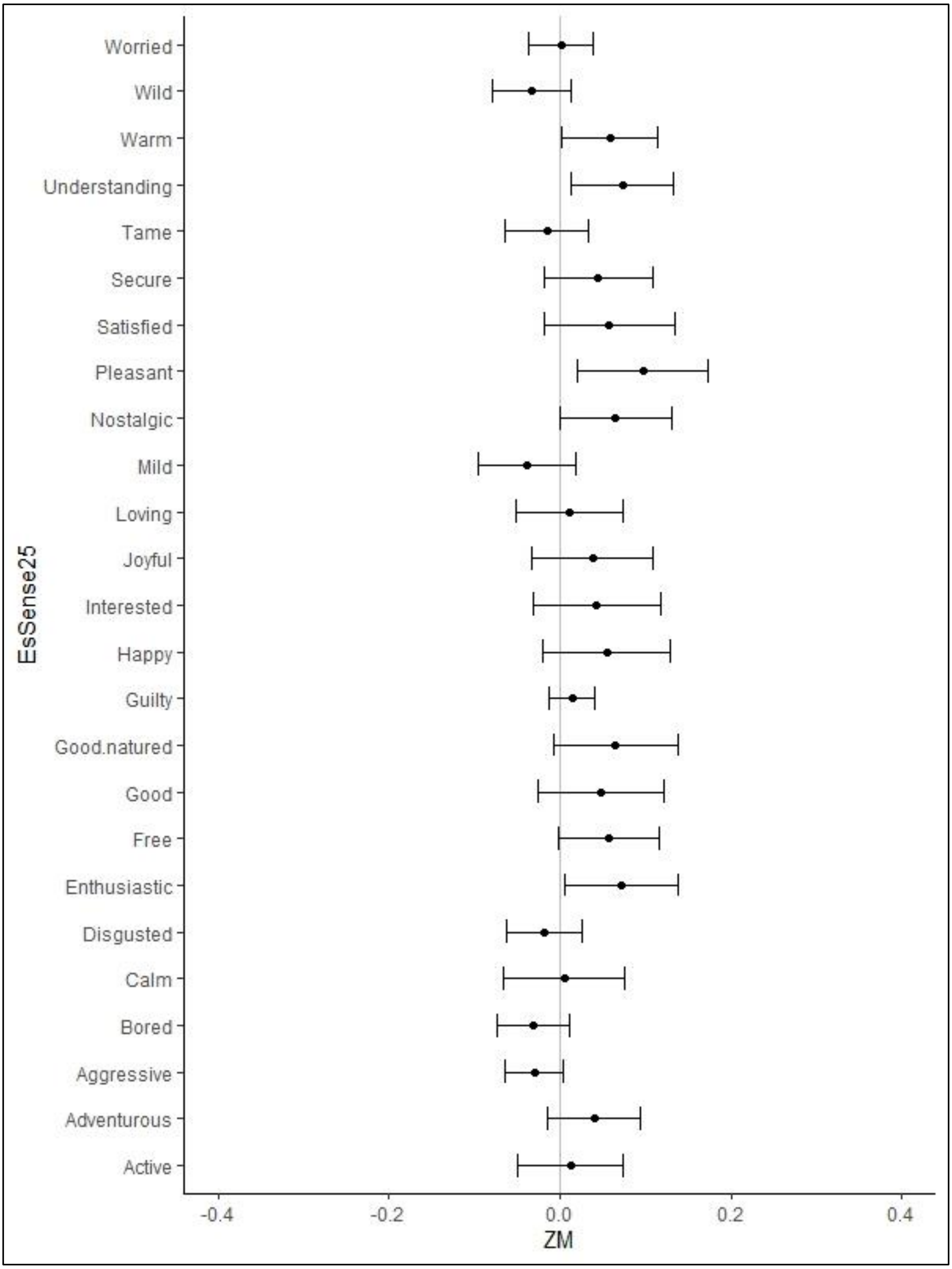

Figure 24: Fixed-effect plot showing the predictive ability of zygomaticus activity for the EsSense 25 profile for milk samples.

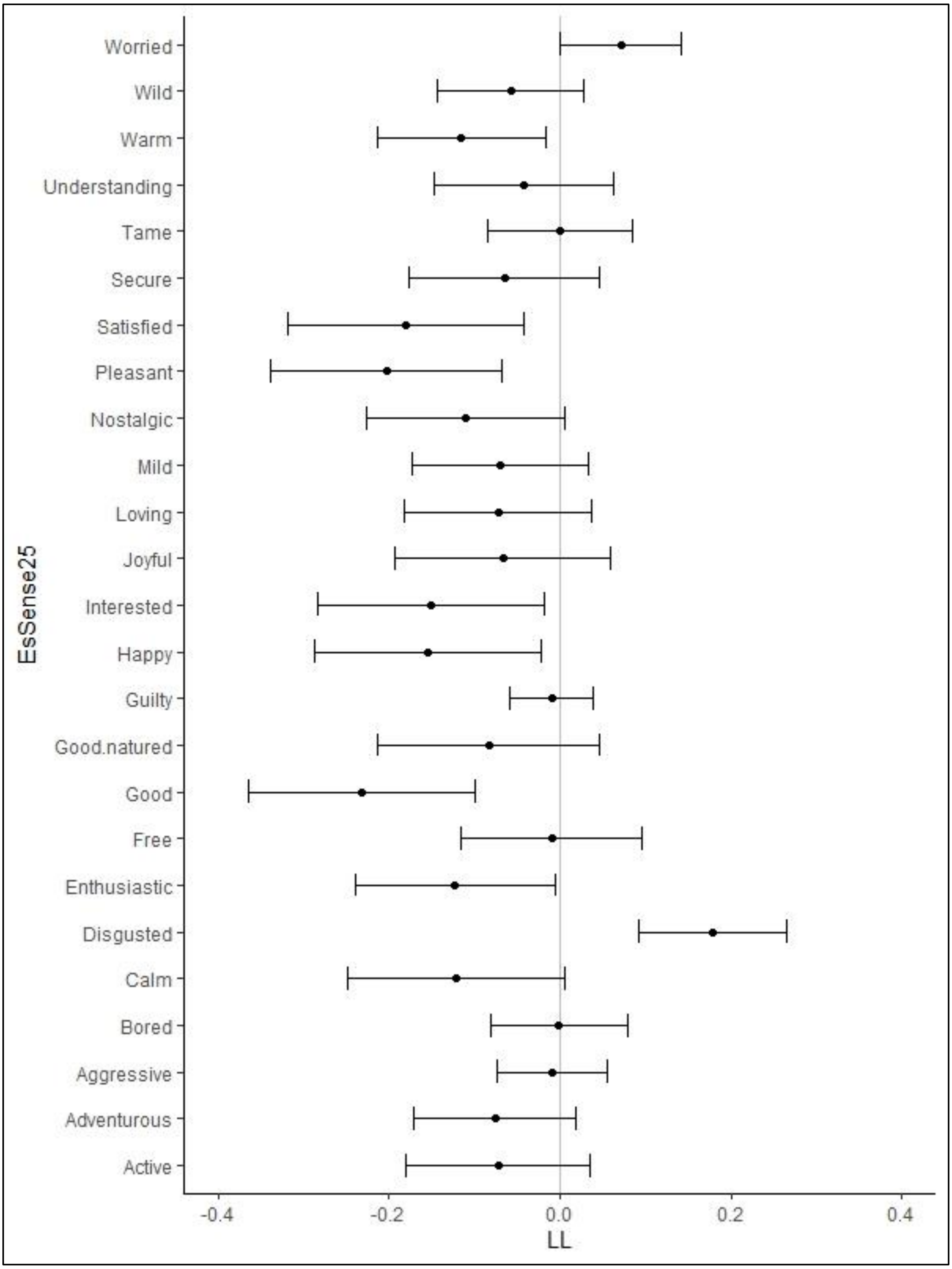

Figure 25: Fixed-effect plot showing the predictive ability of levator activity for the EsSense 25 profile for milk samples.

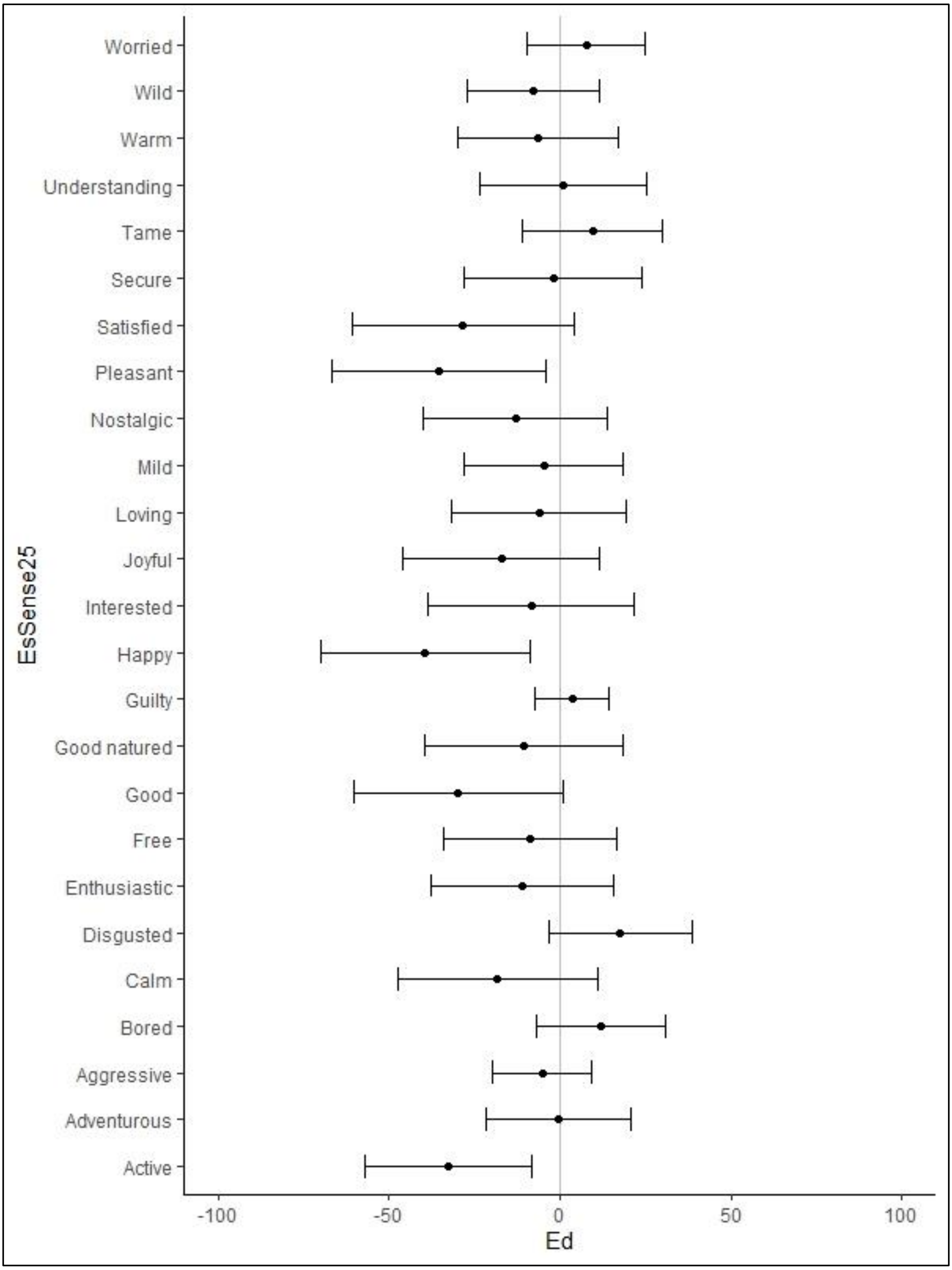

Figure 26: Fixed-effect plot showing the predictive ability of electrodermal activity for the EsSense 25 profile for milk samples.

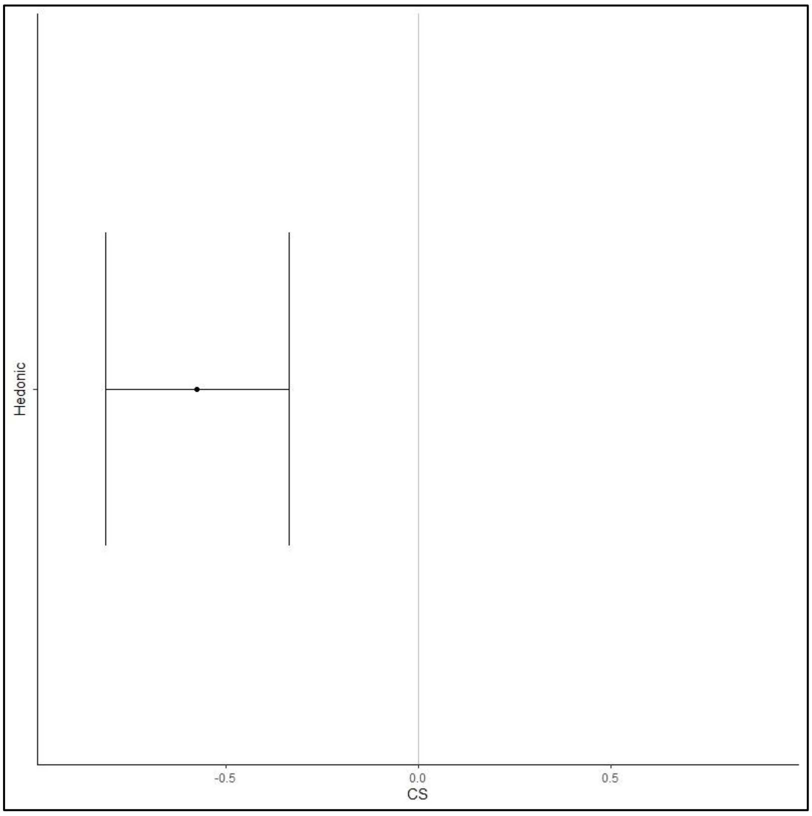

Figure 27: Fixed-effect plot showing the predictive ability of corrugator activity for hedonic liking for milk samples.

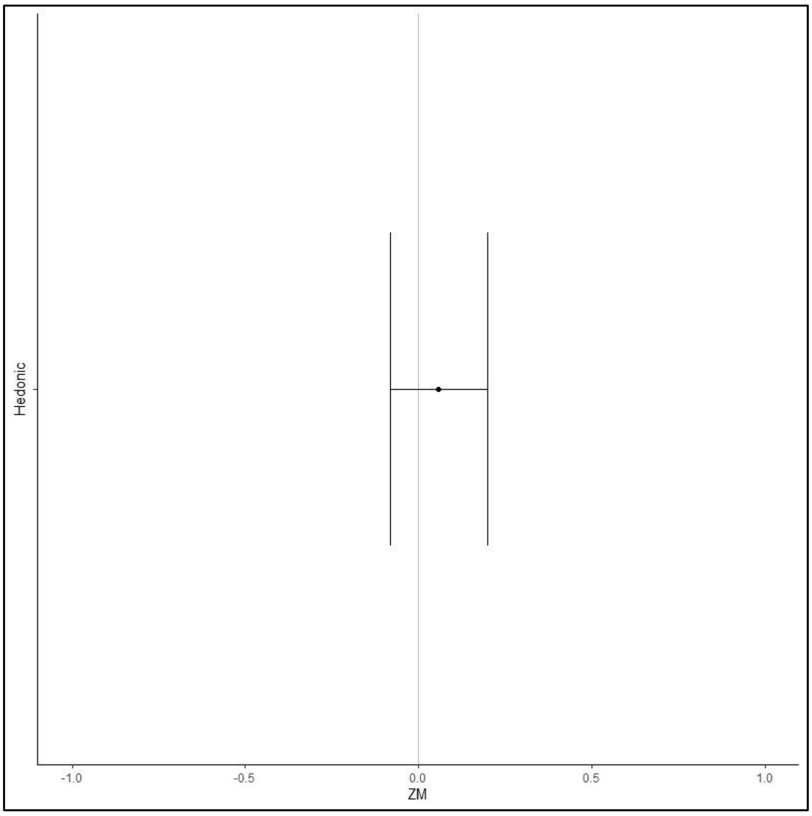

Figure 28: Fixed-effect plot showing the predictive ability of zygomaticus for hedonic liking for milk samples.

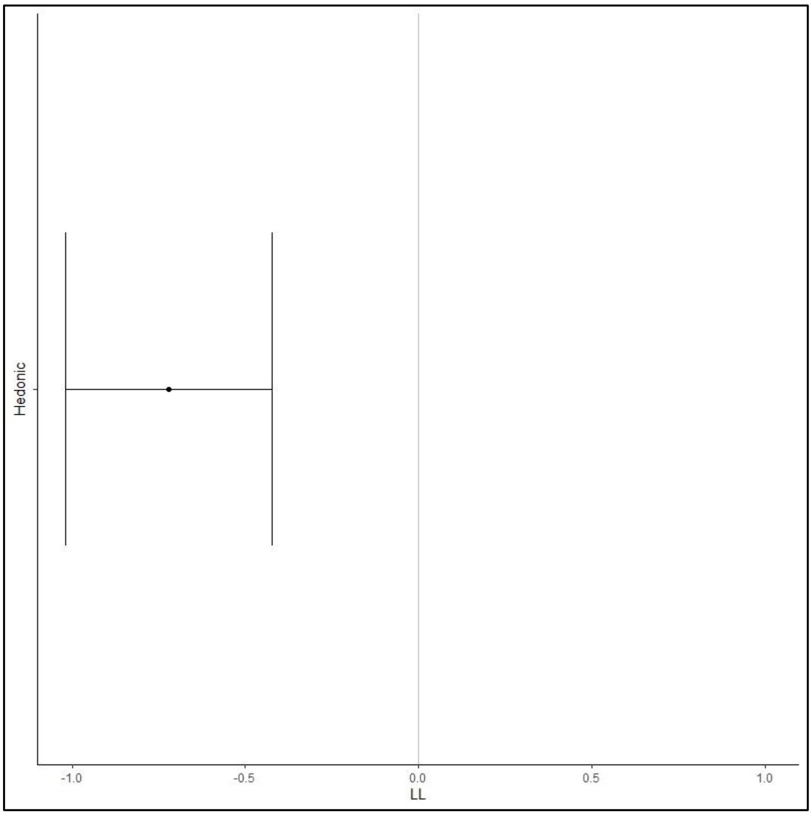

Figure 29: Fixed-effect plot showing the predictive ability of levator for hedonic liking for milk samples.

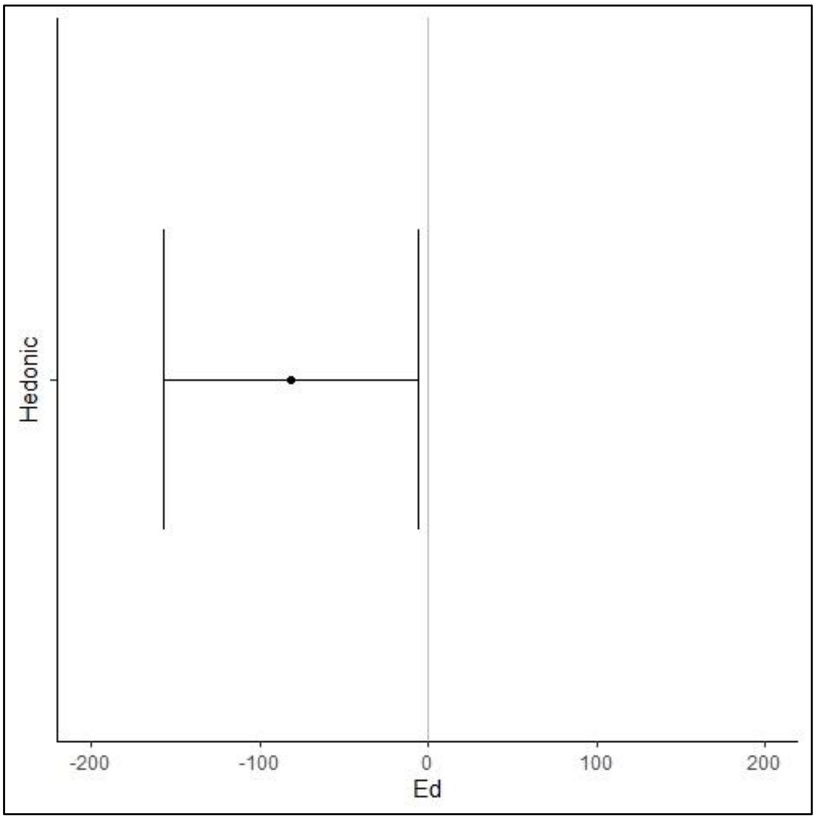

Figure 30: Fixed-effect plot showing the predictive ability of electrodermal activity for hedonic liking for milk samples.

3.2 Assessment of the ability of implicit and explicit measures of emotion and liking to differentiate yoghurt products

The results of three-way ANOVA for each measure of emotion and liking for yoghurt is shown in Table 5 below. There are many statistically significant main effects (particularly of participant), these will be discussed in the sections that follow.

Table 5: p-values for the main effects from the three-way ANOVA for each of the measures of emotion and liking for yoghurt with significant effects in bold.

Measure	Product	Scenario	Participant	Product* Participant	Scenario* Participant	Product* Scenario
Hedonic liking	**0.000**	0.496	**0.012**	**0.000**	0.166	0.142
Active	**0.000**	0.613	**0.000**	**0.000**	**0.000**	0.995
Adventurous	**0.006**	0.058	**0.000**	**0.000**	**0.023**	0.865
Aggressive	**0.000**	0.581	**0.000**	**0.000**	**0.010**	0.917
Bored	**0.000**	0.123	**0.000**	**0.003**	**0.000**	0.210
Calm	**0.000**	0.596	**0.000**	**0.000**	**0.000**	**0.014**
Disgusted	**0.000**	0.892	**0.042**	**0.000**	**0.000**	0.462
Enthusiastic	**0.000**	0.867	**0.000**	**0.000**	**0.000**	0.782
Free	**0.000**	0.557	**0.000**	**0.000**	**0.000**	0.717
Good	**0.000**	0.769	**0.000**	**0.000**	0.069	0.606
Good natured	**0.000**	0.296	**0.000**	**0.000**	**0.000**	0.846
Guilty	0.063	0.706	**0.000**	**0.000**	**0.013**	0.234
Happy	**0.000**	0.587	**0.000**	**0.000**	**0.005**	0.213
Interested	**0.000**	0.825	**0.000**	**0.000**	**0.017**	0.957
Joyful	**0.000**	0.909	**0.000**	**0.000**	**0.000**	0.830
Loving	**0.000**	0.260	**0.000**	**0.000**	**0.000**	0.632
Mild	0.115	0.508	**0.000**	**0.017**	**0.000**	0.544
Nostalgic	**0.000**	0.783	**0.000**	**0.000**	**0.018**	0.056
Pleasant	**0.000**	0.427	**0.000**	**0.000**	**0.003**	0.657
Satisfied	**0.000**	0.687	**0.001**	**0.000**	**0.010**	0.553
Secure	**0.000**	0.486	**0.000**	**0.000**	**0.000**	0.091
Tame	0.127	0.951	**0.000**	0.251	**0.000**	0.526
Understanding	**0.000**	0.582	**0.000**	**0.000**	**0.000**	0.787
Warm	**0.000**	0.880	**0.000**	**0.000**	**0.000**	0.428
Wild	0.139	0.304	**0.000**	**0.000**	**0.000**	0.249
Worried	**0.000**	0.410	**0.000**	**0.000**	0.255	0.994
corrugator	**0.000**	0.657	**0.000**	**0.000**	**0.000**	0.751
zygomaticus	**0.000**	0.210	**0.000**	**0.027**	**0.000**	0.598
levator	**0.000**	0.873	**0.000**	**0.000**	**0.000**	0.225
Phasic EDA	0.141	0.224	**0.001**	0.958	0.420	0.156

3.2.1 Hedonic liking of yoghurt products

A significant product effect on yoghurt liking was evident, with a Tukey post-hoc test identifying four subsets (all products were differentiated from each other except for Fresh n' Fruity Reduced Sugar and Gopala Natural), however there was also a significant participant effect and product*participant interaction ($p<0.05$). Subsequent cluster analysis revealed four participant clusters (cluster 1: n=21, cluster 2: n=14, cluster 3: n=11, and cluster 4: n=13) with different liking patterns. Three of the four clusters however liked the kefir least. Cluster 1 rated liking high for all products except for Fresh n' Fruity Reduced Sugar Natural and liked all of the products except for The Collective Kefir. Cluster 2 followed a similar pattern as cluster 1 but did not like the Gopala Natural yoghurt. Cluster 3 was characterised by disliking of Puhoi Valley, and cluster 4 (n=13) was differentiated by also disliking the Fresh n' Fruity Greek-style Natural.

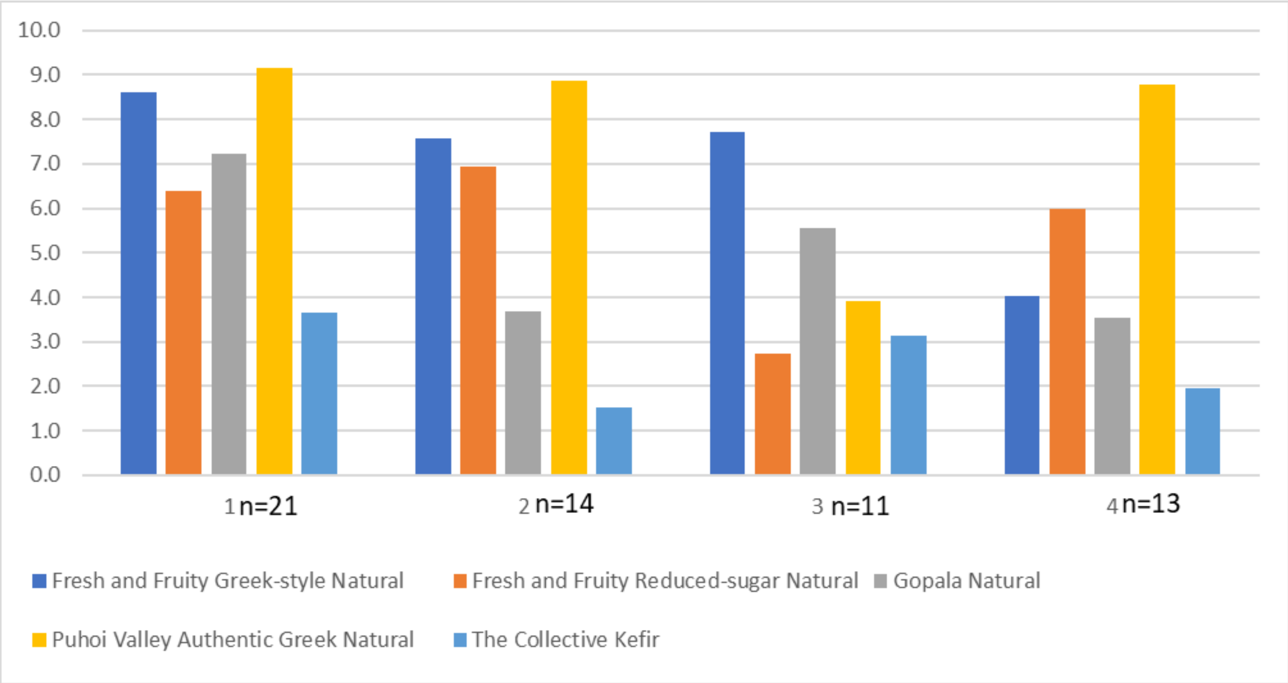

Figure 31: Mean rating of hedonic liking (on a 0-10 scale) for each yoghurt product for the four clusters of participants grouped by their liking ratings for yoghurt.

3.2.2 Ratings of EsSense 25 lexicon terms for yoghurt products

Overall, the emotional engagement of participants with 'yoghurt' was low, with 14 emotions (Active, Adventurous, Aggressive, Bored, Free, Guilty, Mild, Nostalgic, Secure, Tame, Understanding, Warm, Wild, Worried) receiving mean scores between 0 and 1, and eight (Calm, Disgusted, Enthusiastic, Good natured, Interested, Joyful, Loving, Pleasant) between 0 and 2. Only Good, Happy and Satisfied scored between 0 and 3, however most emotion terms had a significant product effect (22 of 25) and all 25 terms had a significant participant effect ($p<0.05$), as shown in Table 5.

Without segmentation, participants rated Active differently to hedonic liking, with similar ratings of Active for Fresh n' Fruity Reduced Sugar Natural, Gopala Natural, and The Collective Kefir despite a 3 to 4-point difference in hedonic liking.

Cluster analyses on each lexicon term found that whilst a term might be statistically significant, it was not necessarily used by many participants, for example Bored (Figure 32).

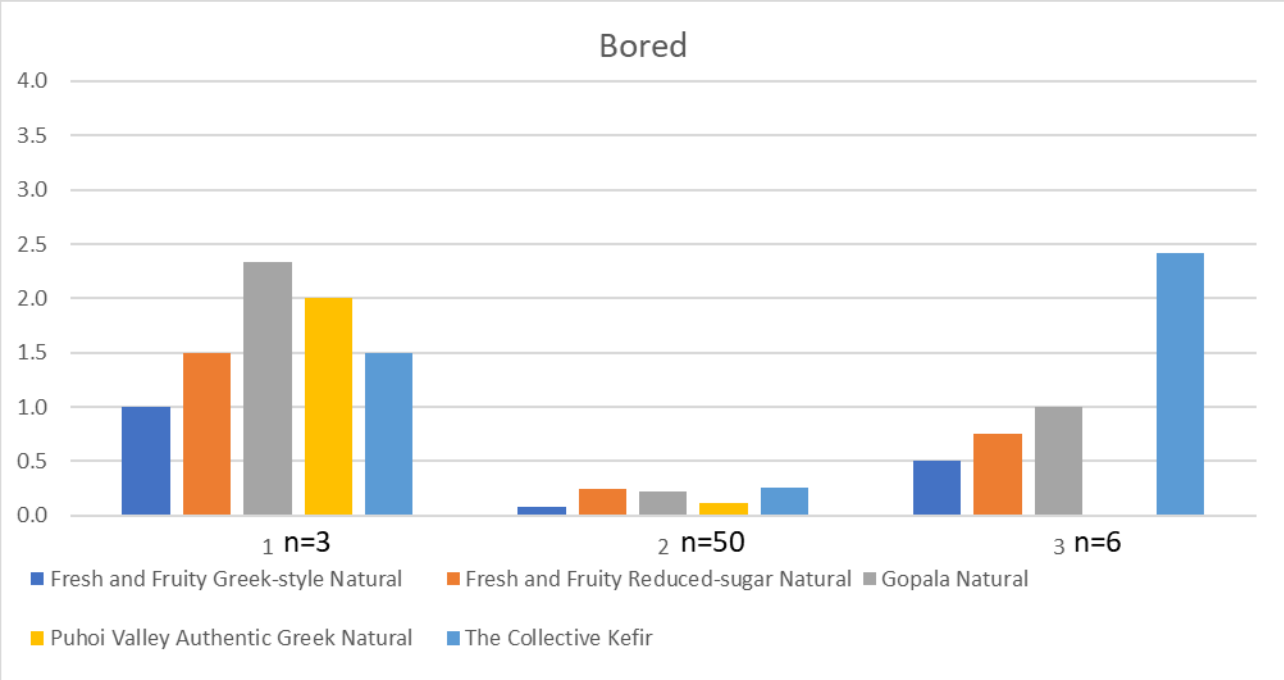

Figure 32: Mean rating of Bored (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Bored' for yoghurt.

For each term there was a cluster of participants who generally scored emotions lower than "1" ("slightly") for all products with the cluster size varying by emotion (n= 21 to 50), Figure A24 to Figure A47 in Appendix 4. Terms were considered 'relevant to most participants' when more than 50% of participants rated the term above "1" for at least one product. There were 13 terms which had a significant effect of product ($p<0.05$) but were not relevant to most participants: Adventurous, Aggressive, Bored, Calm, Enthusiastic, Free, Guilty, Joyful, Nostalgic, Secure, Understanding, Warm and Worried (Table 6).

Table 6: Groups of EsSense 25 terms based on significant main effect of product (p<0.05) and relevance for yoghurt. Terms were considered relevant when at least 50% of participants were in clusters that had mean scores > 1 ("slightly") for at least 1 product.

	Not significant	Significant
Not relevant to most participants	Mild, Tame, Wild	Adventurous, Aggressive, Bored, Calm, Enthusiastic, Free, Guilty, Joyful, Nostalgic, Secure, Understanding, Warm, Worried
Relevant to most participants		Active, Disgusted, Good, Good natured, Happy, Interested, Loving, Pleasant, Satisfied

Nine terms were considered relevant and had a significant main effect of product (p<0.05) for yoghurt products (Active, Disgusted, Good, Good natured, Happy, Interested, Loving, Pleasant and Satisfied).

3.2.3 Facial muscle movements in response to consuming yoghurt products

For *corrugator* activity, there was a significant effect of product and participant on the muscle activity, but also significant interactions of product*participant and scenario*participant (p<0.05). Overall, *corrugator* activity could differentiate The Collective Kefir from the other products, with a Tukey post-hoc test finding it in a separate subset to the remaining products. However, this does not consider the participant main effect and product*participant interactions which indicate that the effect of product on the *corrugator* activity may depend on the participant. This is evident in the three clusters obtained from the cluster analysis (Figure 33), which differ in both intensity of *corrugator* activity and the pattern of *corrugator* activity across the products. However, The Collective Kefir always had the greatest *corrugator* activity across the clusters, with cluster 2 having the largest different between this and the other products. Cluster 2 was also different in that the Puhoi product had low *corrugator* activity compared to the other products.

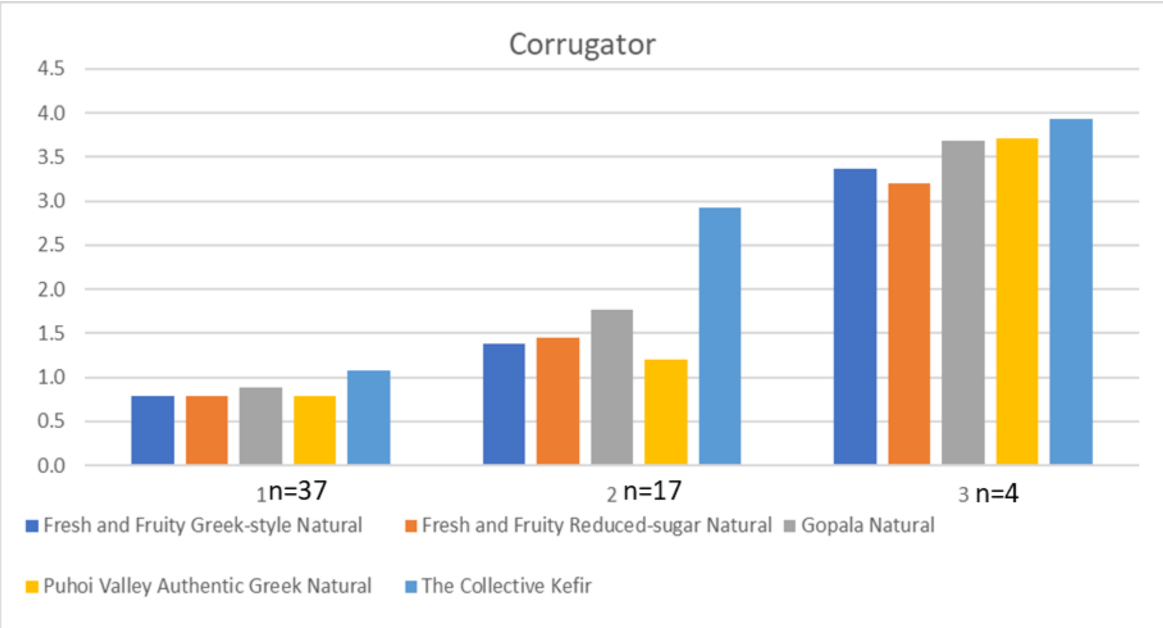

Figure 33: Mean corrugator activity (percentage of the maximum) for each yoghurt product for the three clusters (1-3) of participants as grouped by their corrugator activity.

For *zygomaticus* activity, there was a significant effect of product and participant on the muscle activity, and significant interactions of product*participant and scenario*participant (p<0.05). A Tukey post-hoc identified two subsets with the Fresh n' Fruity Greek-style and Puhoi Valley products in a separate subset from the remainder. Due to the participant main effect and interaction with product, the data was looked at more closely using a cluster analysis. Three of the four clusters (clusters 1, 2 and 4) follow the same pattern of *corrugator* with higher activity for the Fresh n' Fruity Greek-style and Puhoi Valley products (Figure 34). These clusters are differentiated by the level of *zygomaticus* activity across all the products, with cluster 1 (n=27) having the lowest activity, followed by cluster 2 (n=21) and then cluster 4 (n=9). Cluster 3 had the highest *zygomaticus* activity and a different pattern, with increased muscle movement for the Fresh n' Fruity Greek-style and The Collective Kefir, however this cluster contained only two participants.

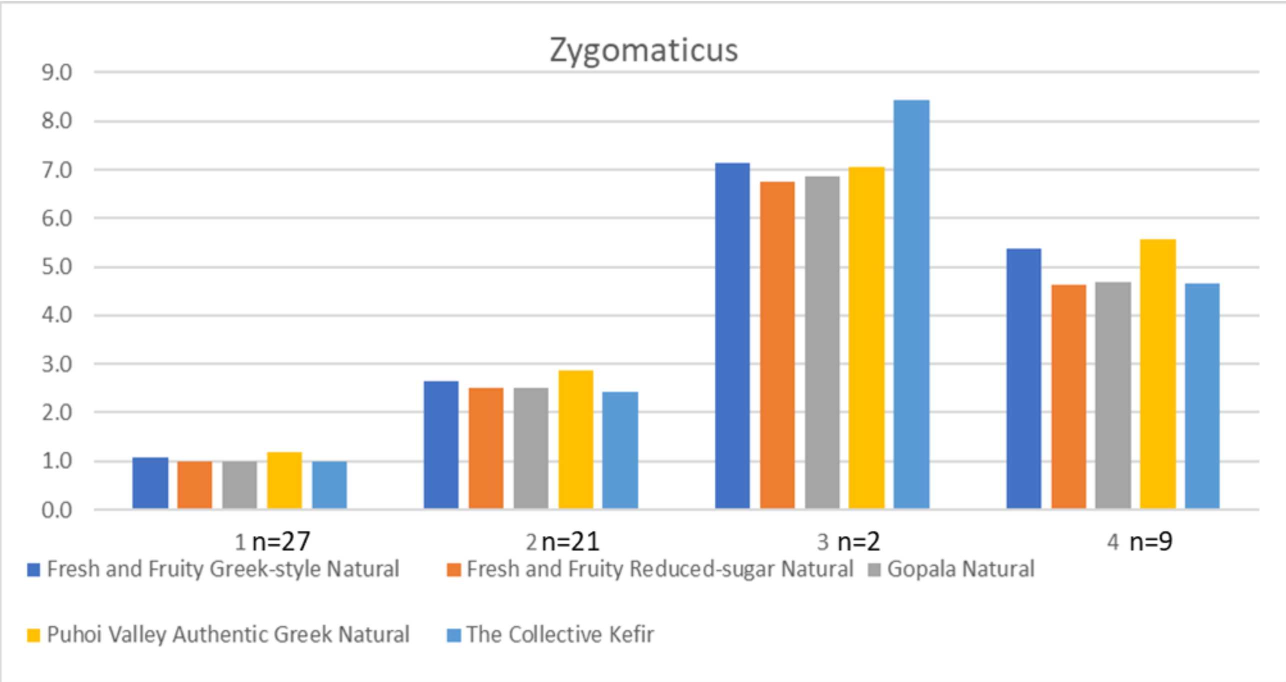

Figure 34:. Mean zygomaticus activity (percentage of the maximum) for each yoghurt product for the four clusters (1-4) of participants as grouped by their zygomaticus activity.

For *levator* activity, there was a significant effect of product and participant on the muscle activity ($p<0.05$) and a Tukey post-hoc test found that there were three subsets (with significantly less *levator* activity associated with the Fresh n' Fruity Reduced-sugar product than The Collective Kefir and the other products falling between these two). There was also significant interactions of product*participant and scenario*participant, indicating that the relationship between *levator* activity and product varied between participants, which was confirmed with a cluster analysis identifying four clusters (Figure 35). The only product differentiation consistent across all four clusters was that the Fresh n' Fruity Reduced-sugar Natural yoghurt elicited lower *levator* activity than The Collective Kefir. Participants in cluster 1 had low activity although this was higher for the Kefir. The pattern of clusters 2 and 3 were similar except *levator* activity was much lower for cluster 2. The person's muscle activity in cluster 4 was particularly strong to Fresh n Fruity Greek and Gopala, then kefir showing quite a different pattern to the other participants indicating that they may be an outlier.

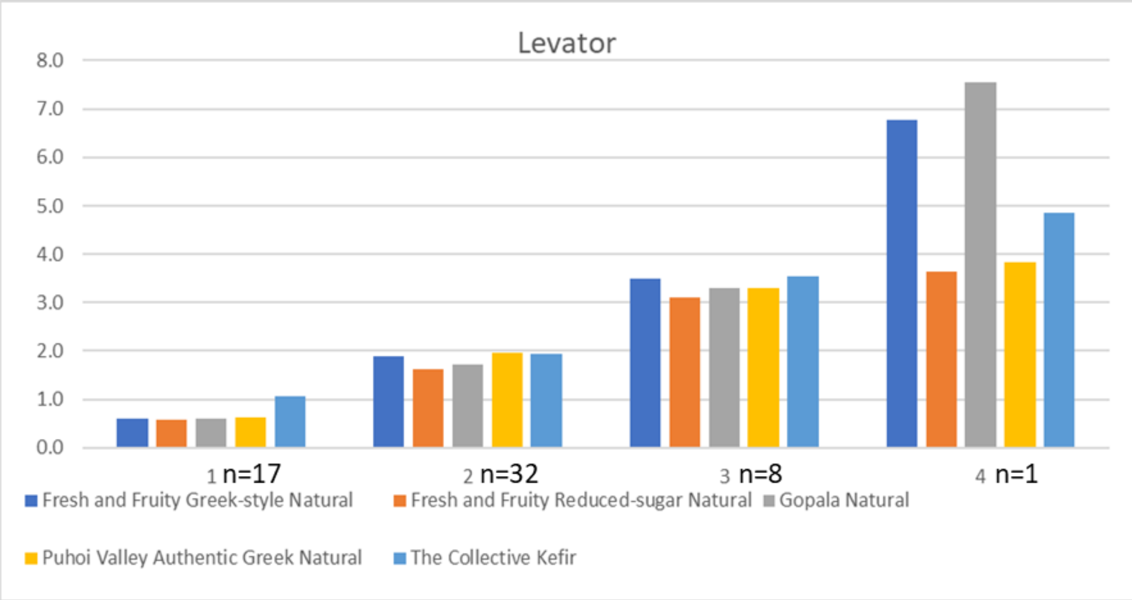

Figure 35: Mean levator activity (percentage of the maximum) for each yoghurt product for the four clusters (1-4) of participants as grouped by their levator activity.

3.2.4 Phasic EDA in response to consuming yoghurt products

Participant was the only significant factor ($p<0.05$) in the phasic EDA activity ANOVA for the yoghurt products (Table 5). Cluster analysis revealed four clusters of participants (Figure 36), with the majority showing little phasic activity for any product (cluster 1: n= 46).

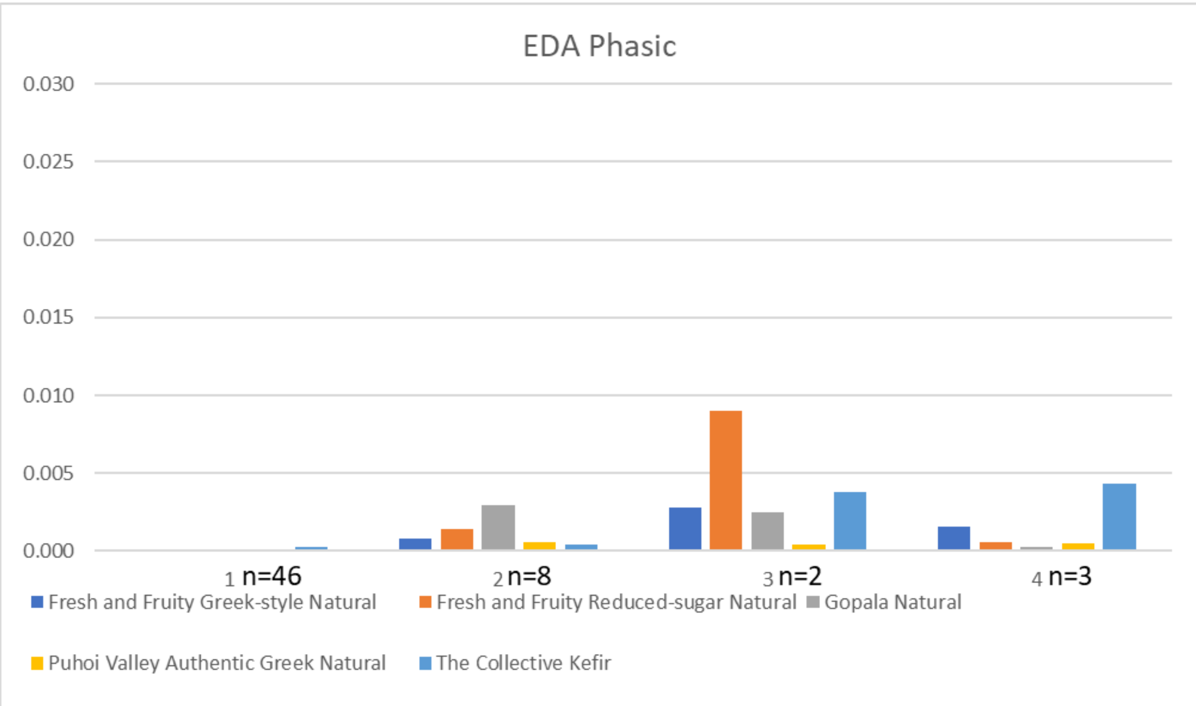

Figure 36: Mean electrodermal activity (in micro siemens) for each yoghurt product for the four clusters (1-4) of participants as grouped by their electrodermal activity.

The remaining 13 were spread over three clusters (cluster 2: n=8, cluster 3: n=2, cluster 4: n=3) with varying patterns of activity over the products. The eight participants in cluster 2 had increased EDA for the Gopala product, and also the Fresh n' Fruity Reduced-sugar. Cluster 3 (n=2) had the highest EDA activity for the Fresh n' Fruity Reduced-sugar (0.009) and Fresh n' Fruity Greek-style (0.003), however the activity for the other products were similar to cluster 2 (Gopala) and cluster 3 (The Collective Kefir). Cluster 4 (n=3) had increased EDA for The Collective Kefir and the Fresh n' Fruity Greek-style, however all values were less than 0.005.

3.2.5 Correlations and predictive ability of implicit and explicit measures of emotion and liking for yoghurt products

There were no significant correlations between the implicit EDA and EMG measures and explicit measures of emotional response (EsSense25) and liking (Figure 37 , r = -0.3 to 0.2).

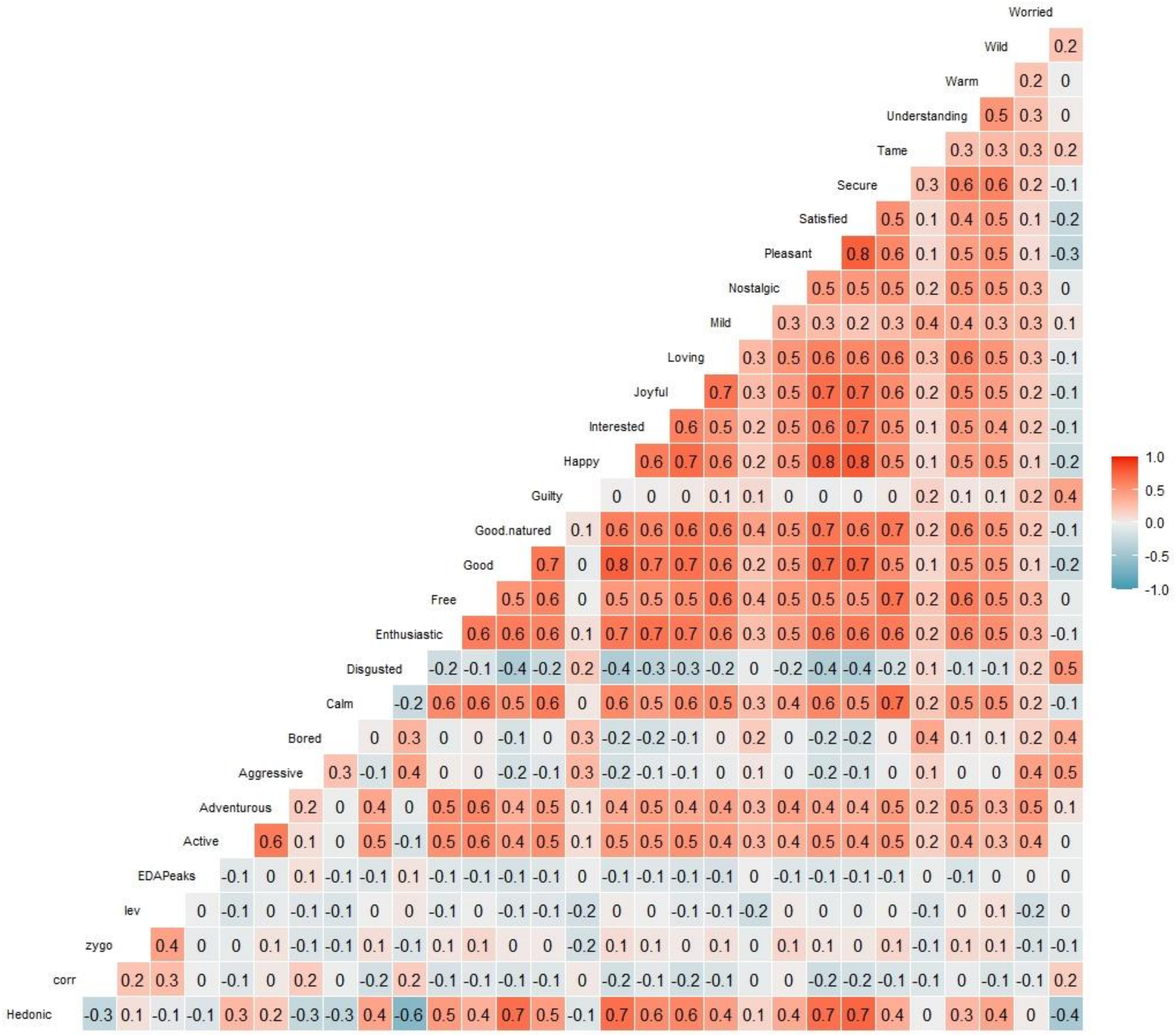

Figure 37: Pearson correlation coefficients for the hedonic liking, implicit and explicit emotion measures for yoghurt samples.

However, there were some positive correlations between hedonic liking and specific EsSense 25 terms. Good, Happy, Pleasant, and Satisfied had correlation coefficients of 0.7, and Interested and Joyful had a correlation coefficient of 0.6. These lexicon terms correlated with each other, with coefficients ranging from 0.6 to 0.8. There was also a negative correlation between hedonic liking and Disgusted with a correlation coefficient of -0.6.

Interestingly, there was a positive correlation between Calm and Secure (0.6), and between Secure and Wild (0.6), however was no correlation between Calm and Wild (0.2). This may be due to only a small number of participants using the terms rather than the whole population.

Despite any correlation between implicit and explicit emotion measures (Figure 37), when considered independently, the muscles measured with EMG had the ability to predict some terms in the EsSense 25 lexicon and liking. This is shown in the fixed-effect plots (Figure 38 to Figure 45), where the change in score for each lexicon term (or hedonic liking) when a 1-point increase in muscle activity or EDA occurs. The implicit measure is considered a predictor of the explicit measure when the 95% confidence interval does not overlap with the centre line. The corrugator could predict the most terms, with a 1-point change in *corrugator* activity predicting an increase in the rating of three terms (Aggressive, Disgusted and Worried) and a decrease in rating for 16 terms: Calm, Enthusiastic, Free, Good, Good natured, Happy, Interested, Joyful, Loving, Mild, Nostalgic, Pleasant, Satisfied, Secure, Understanding and Warm (Figure 38). Five EsSense 25 terms were predicted by the *zygomaticus* (Calm, Enthusiastic, Good natured, Secure and Warm), with a 1-point increase in zygomaticus activity predicting an increase in rating of around 0.1 points. Only two terms were predicted by the *levator*: Good and Interested. Both terms have similar point estimates (approximately -0.1), however the confidence interval for Good is further from the centre line of the plot than the one for Interested. Whilst the *levator* had the ability to predict fewer EsSense 25 terms than the *zygomaticus*, the *levator* was able to predict liking whilst the *zygomaticus* could not. The *corrugator* could also predict liking, and was a better predictor than the *levator*, with a 1-point change in muscle activity predicting a 0.6-point and 0.3-point decrease in liking rating for the *corrugator* and *levator* respectively.

EDA had no predictive ability for any EsSense 25 term nor liking) for the yoghurt products. All confidence intervals crossing the centre line of the plots in Figure 41 and Figure 45.

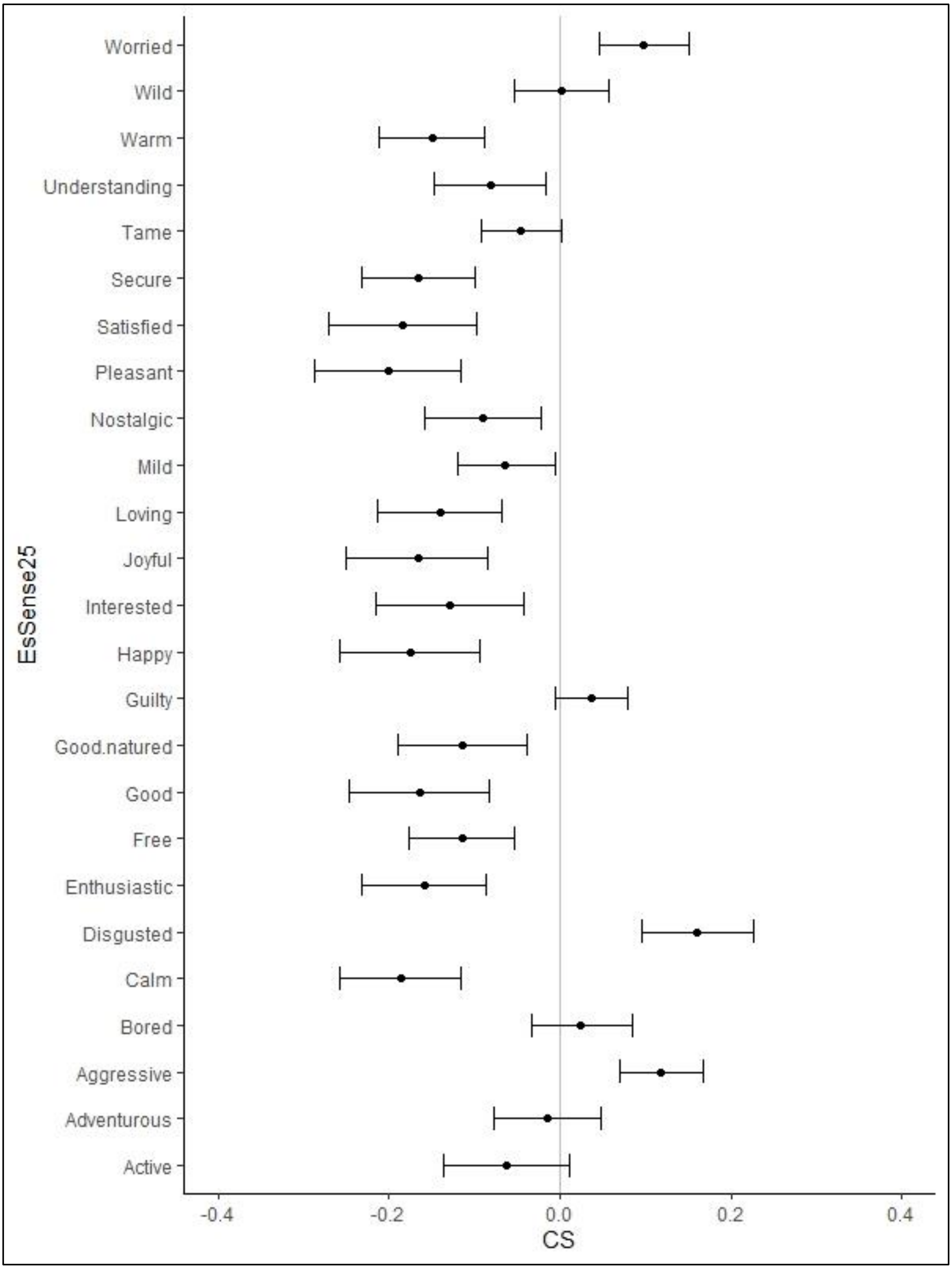

Figure 38: Fixed-effect plot showing the predictive ability of corrugator activity for the EsSense 25 profile for yoghurt samples.

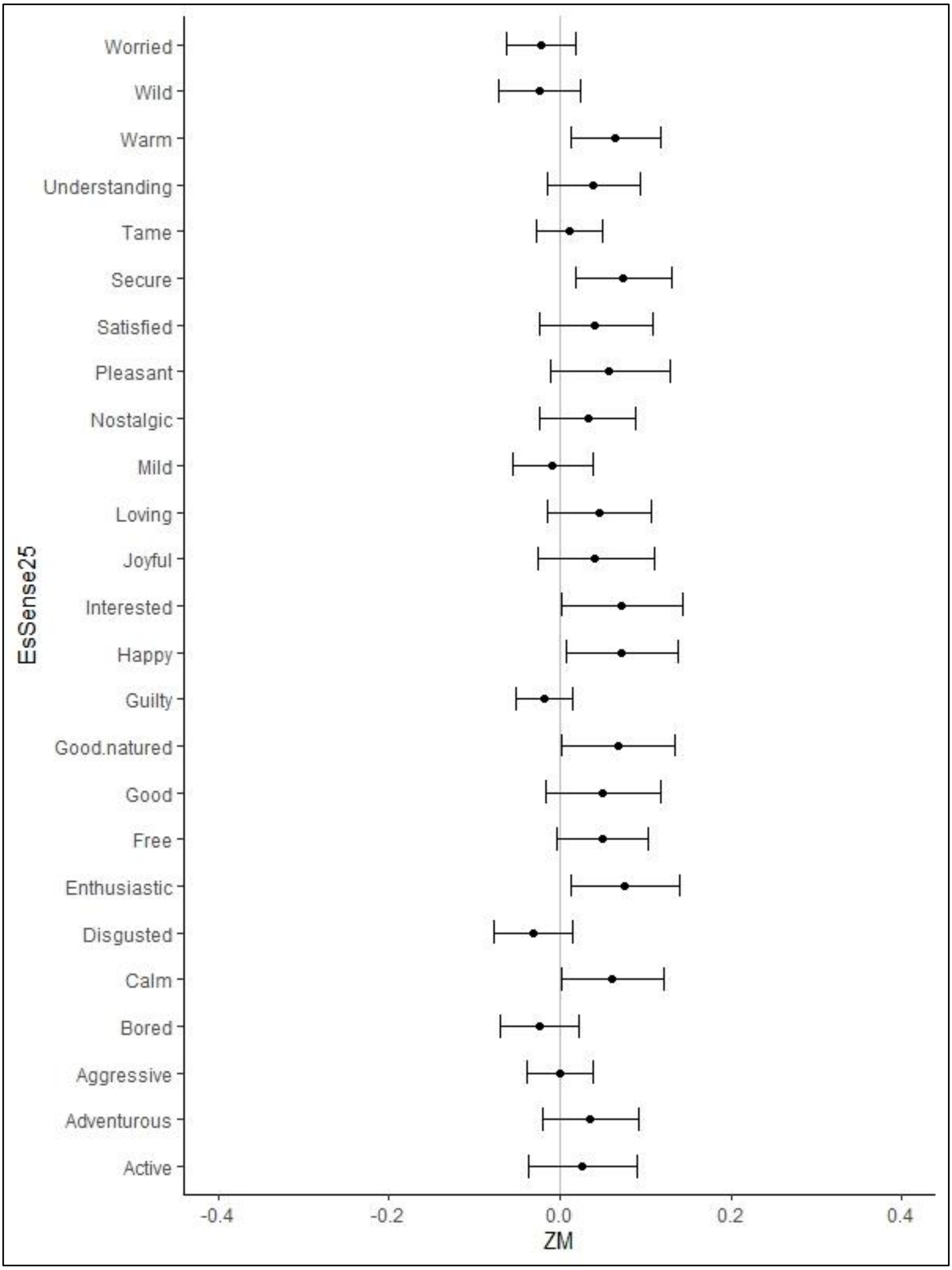

Figure 39: Fixed-effect plot showing the predictive ability of zygomaticus activity for the EsSense 25 profile for yoghurt samples.

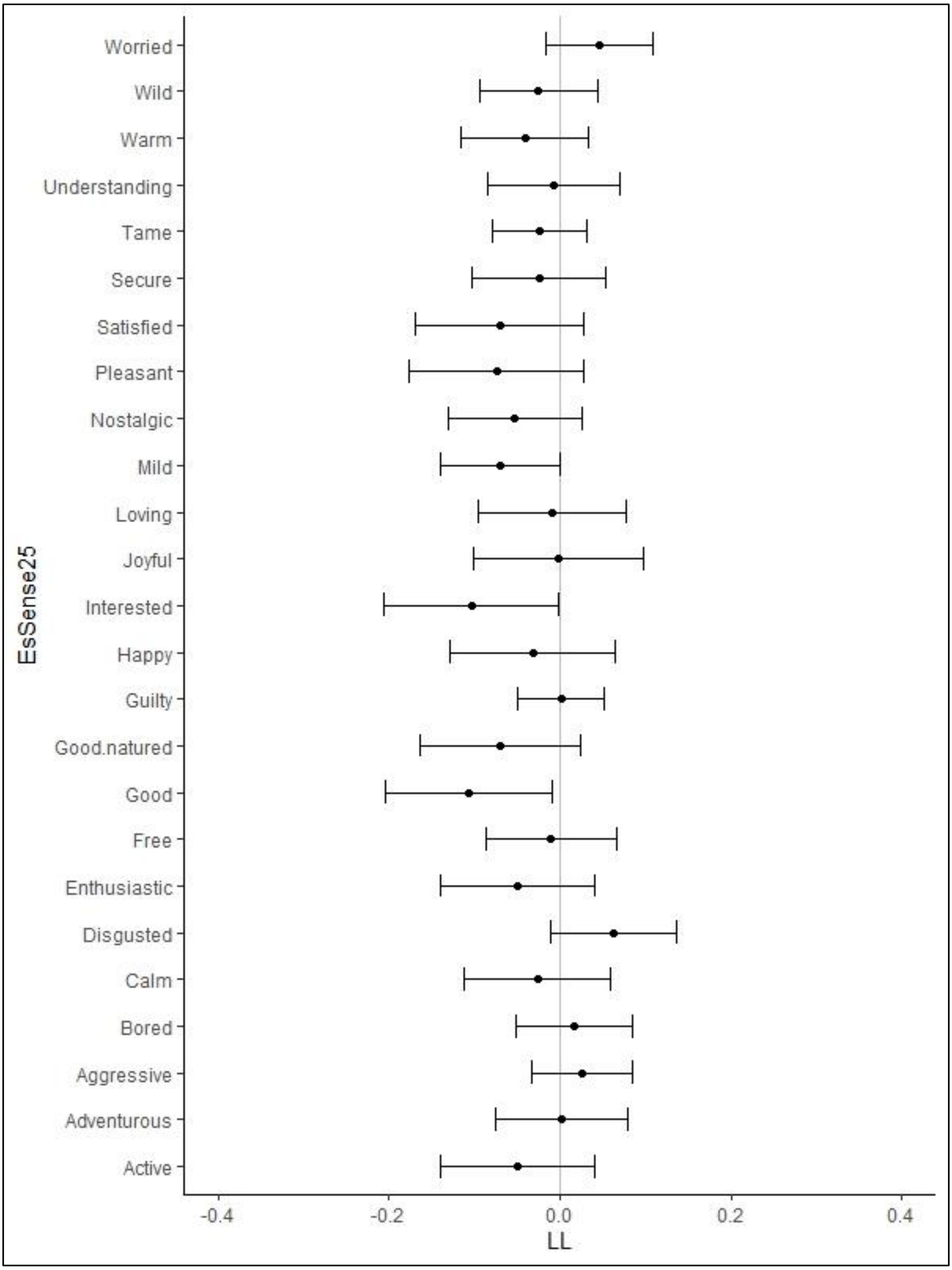

Figure 40: Fixed-effect plot showing the predictive ability of levator activity for the EsSense 25 profile for yoghurt samples.

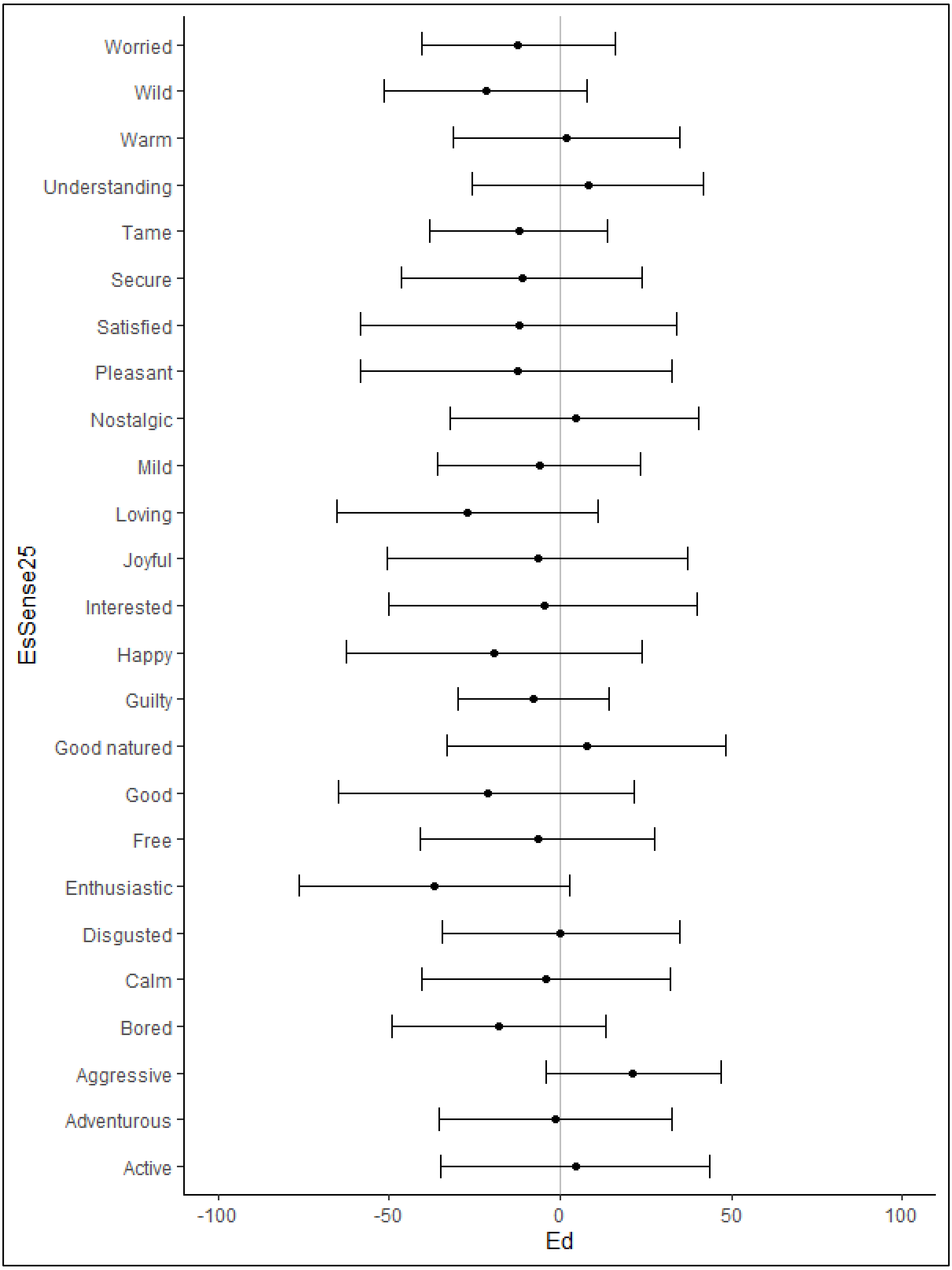

Figure 41: Fixed-effect plot showing the predictive ability of electrodermal activity for the EsSense 25 profile for yoghurt samples.

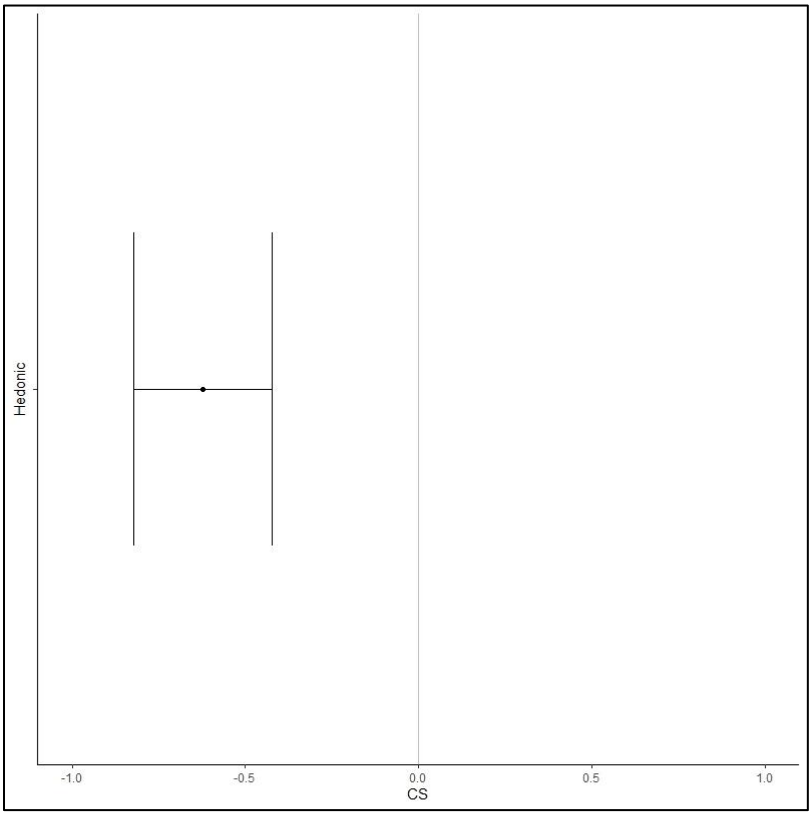

Figure 42: Fixed-effect plot showing the predictive ability of corrugator activity for hedonic liking of yoghurt samples.

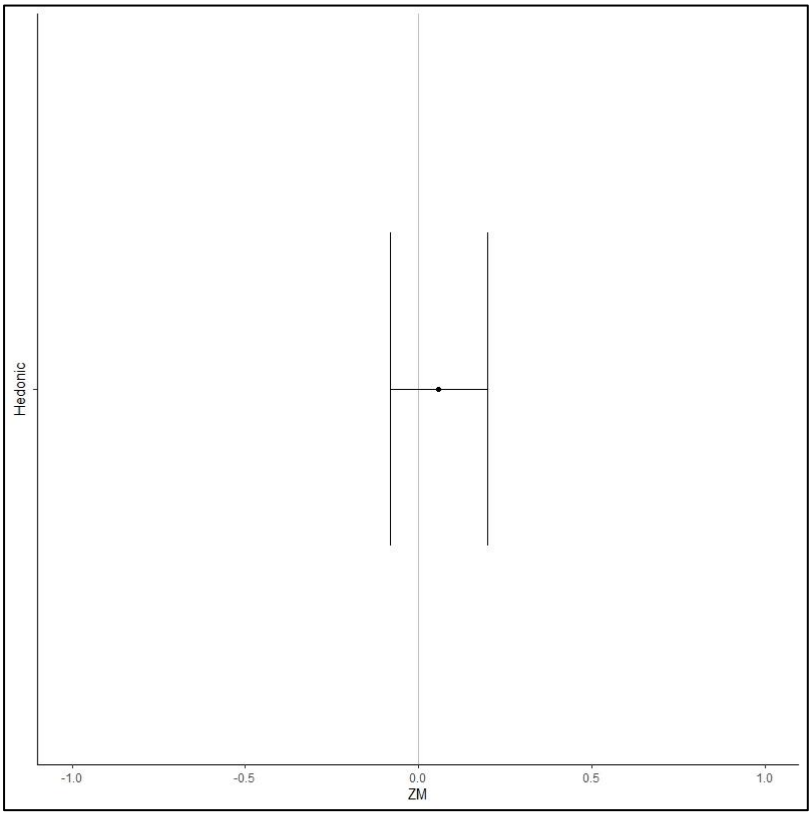

Figure 43: Fixed-effect plot showing the predictive ability of zygomaticus activity for hedonic liking of yoghurt samples.

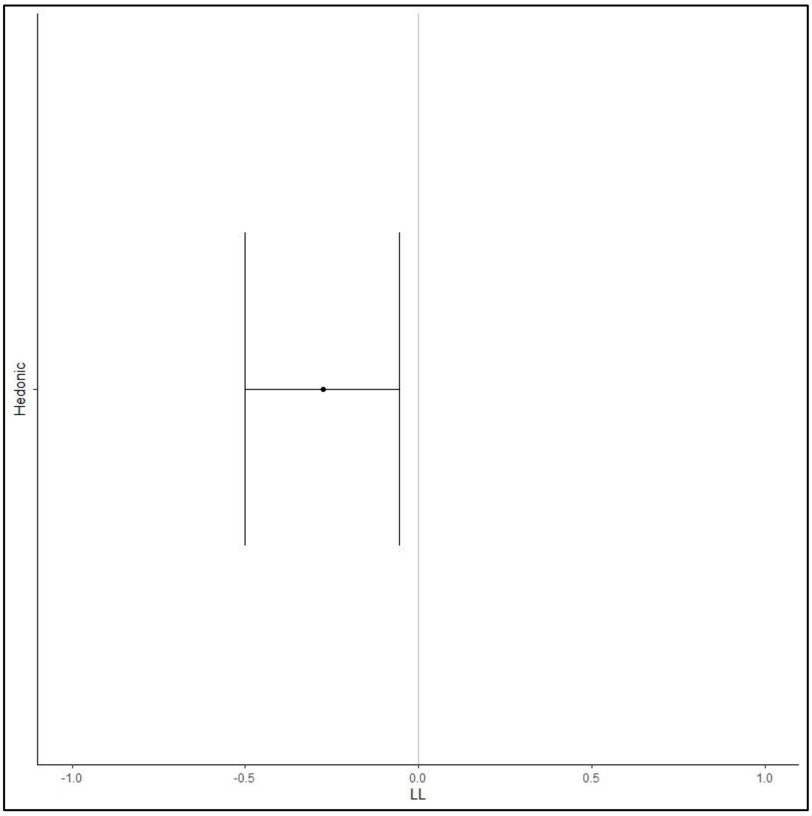

Figure 44: Fixed-effect plot showing the predictive ability of levator activity for hedonic liking of yoghurt samples.

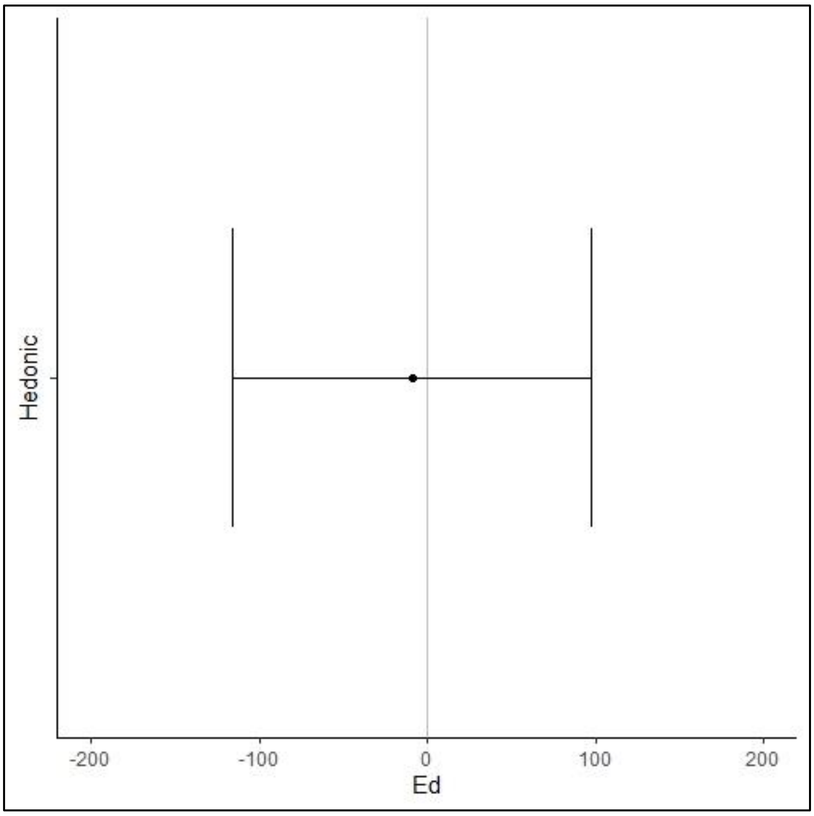

Figure 45: Fixed-effect plot showing the predictive ability of electrodermal activity for hedonic liking of yoghurt samples.

4 DISCUSSION

The aim of this study was to investigate the ability of specific explicit and implicit measures of emotional response and liking to discriminate milk and yoghurt products. Specifically, the study was guided by the following research questions:

- Can milk and yoghurt products be differentiated by selected explicit measures of emotion and liking?
- Can the milk and yoghurt products be differentiated by selected implicit measures of emotion?
- Do the implicit and explicit measures of emotional response and liking correlate?
- Do the implicit and explicit measures of emotional response and liking differentiate milk and yoghurt to different extents?
- Are the selected implicit and explicit measures impacted by the use of an individually composed written evoked scenario?

4.1 CAN MILK AND YOGHURT PRODUCTS BE DIFFERENTIATED BY EXPLICIT MEASURES OF EMOTION AND LIKING?

When looking at the whole set of participants, hedonic liking scores could differentiate products within the milk and yoghurt product sets, however whilst liking could differentiate four of five yoghurt products it only separated the two least liked milk products from the remaining three. This separation of the more extreme products within a product set with small differences in sensory characteristics is consistent with what Rocha et al. (2019) found with Lemon Verbena tea. Nevertheless, whilst the mean liking scores could differentiate products the pattern of liking was not consistent across all participants, with participants grouped into three clusters based on their liking of milk and four clusters for yoghurt. Within the clusters, the patterns of liking were more evident, for example, cluster 3 for yoghurt disliked the Fresh n' Fruity Reduced-sugar and the Puhoi Valley products whereas the other clusters liked these products. The Reduced-sugar product was sweetened with an artificial sweetener and the Puhoi Valley product had added sugar, whilst the other products were unsweetened, indicating that this cluster was driven by their dislike of sweet yoghurts. The implication of this is that there are groups within the population that have different sensory drivers of liking. These should be considered when developing new products, as creating a product based on average liking scores may end up satisfying few consumers (Schilling and Coggins 2007). However, running a cluster analysis alone does not provide these insights, additional statistical analyses (such as ANOVA) within clusters should be run. Whether meaningful conclusions can be drawn from these statistics depends in part on the number of participants in each cluster (Nguyen and Wismer 2019), so typically consumer acceptance studies record responses from more than 100 participants.

Less than half of the emotion terms in the EsSense 25 lexicon were able to differentiate between products and considered relevant to most participants, with 10 for milk and nine for yoghurt. Notably, the ratings of terms were dependent on the individual participant, with cluster analysis showing that some terms were only important to small groups of participants which was driving the significant difference in the overall mean scores. Clustering of lexicon terms based on how they are rated by participants has been used to segment consumers based on the reduced number of emotion categories (Low et al. 2021), and could be useful in cases such as these. In addition, collecting data from a larger number of consumers will likely mean that there are more participants in emotionally engaged clusters. The low emotional engagement is potentially a consequence of milk and yoghurt not eliciting strong emotional responses that might be associated with other categories such as chocolate, ice cream and wine, alternatively it may indicate a lack of fit of the emotional lexicon used (Mora, Urdaneta, and Chaya 2018), or both. A product-specific lexicon may improve the emotional engagement of participants, for example the Coffee Drinking Experience (CDE) lexicon created emotional profiles that were more meaningful than those from EsSense® profile for coffee products, (Kanjanakorn and Lee 2017).

Whilst liking could discriminate more products than the EsSense 25 lexicon for both the milk and yoghurt product sets when looking at the data, some emotion terms discriminated between products differently to liking. For example, on average participants rated Bored higher for Anchor Trim Milk than the Gas Station Blue Top Milk contrary to the products having liking scores that were not significantly different. As Bored was not a term that was relevant to most participants, it is possible that the higher rating of Bored for the Anchor Trim Milk came from the cluster of participants which were neutral about the Anchor Trim Milk and liked the Gas Station Blue Top Milk. Alternatively, this could indicate that the lexicon terms are measuring something different to liking, aspect of food reward such as wanting/motivation to eat the food (Berridge 2018), which may add to the ability to predict consumer choice (Gutjar et al. 2015).

4.2 CAN MILK AND YOGHURT PRODUCTS BE DIFFERENTIATED BY IMPLICIT EMOTION MEASURES?

The *corrugator* is typically associated with frowning (Dimberg 1990) and a negative association with hedonic liking, (Nath, Cannon, and Philipp 2019, Sato et al. 2020, Soussignan et al. 2015), however the ability of this method to discriminate tasted food products has not yet been published. In this study, *corrugator* activity was increased for the extreme milk product for 15 participants across three clusters, and the extreme product for yoghurt for all participants. Similarly, the *levator*, associated with nose wrinkling movements (Dimberg 1990), had increased activity for the extreme product for milk for 37 participants across two clusters. The two products at the extremes of like/dislike for yoghurt were able to be discriminated overall, and all clusters of participants showed a similar pattern, with a lack of *levator* activity differentiating the liked product. This aligns with the negative relationship between *levator* activity and liking seen for food images in Nath, Cannon, and Philipp (2019). *Zygomaticus* did not differentiate milk products but did differentiate the two most liked products for yoghurt for 57 of the 59 participants, which is consistent with the finding that *zygomaticus* is positively associated with liked images, (Nath, Cannon, and Philipp 2019).

For all muscles measured using facial EMG, there was a group of participants where the average muscle activity was less than for all products and there was little difference between products. This could be due to absolute low emotional engagement with the products, or participants who are not emotionally expressive through facial expressions. This is supported by the literature, with Ahn and Picard (2014) and Danner et al. (2014) finding that some participants were not emotionally expressive in studies using facial expression analysis, with 50% and 25% falling within this category respectively. Screening potential participants for emotion expressiveness could remove this problem, however the expressive participants may not be a representative sample of the consumer population. Alternatively, collecting data from a larger number of participants may increase the number of expressive participants and provide enough usable data to draw conclusions from. In addition, facial expressions have been suggested to be a form of communication (Soussignan et al. 2015) with certain expressions such as smiling rarely displayed without social interaction (De Wijk et al. 2014). Facial expressions can be affected by the presence of another participant (Nath, Cannon, and Philipp 2020), the researcher (Nath, Cannon, and Philipp 2019) and an animated avatar on the computer screen (Soussignan et al. 2015). This implies that the lack of facial muscle movements in some participants seen in the present study could be due to the isolation of the participant during the task. However, as social context was not included in the present study, no conclusions can be drawn about whether this is the case.

Whilst EDA has been demonstrated to differentiate diluted vinegar from a normal range of beverages, (Kaneko et al. 2019), it was not able to differentiate the products in this study. This could be due to a lack of sensitivity needed to record small differences in response to similar products, as found in Kaneko et al. (2019) where EDA was unable to discriminate between the normal beverages.

4.3 Do the implicit and explicit measures of emotional response and liking correlate?

Assigning meaning to facial expressions is complex because a frown does not always mean that someone is angry, likewise a smile does not always indicate happiness. In addition, the way that facial expressions are used to show emotion depend on factors such as the individual's culture and the situation (Barrett et al. 2019). Unsurprisingly, the correlations between facial EMG and explicit measures found in the literature were small, ranging from 0.04 between *zygomaticus* and liking for food images (Nath, Cannon, and Philipp 2019) to -0.25 between the *corrugator* and liking for flavoured gels (Sato et al. 2020). This is consistent with what was found in the present study, with Pearson correlations showing weak relationships between muscle and liking, the strongest of which was -0.3 between liking and *corrugator* activity for yoghurt. Despite only weak correlations found between implicit and explicit measures, the fixed-effect plots with each facial muscle examined in isolation showed that each muscle was a significant predictor of specific explicit measures. *Corrugator* and *levator* activity could predict hedonic liking with a negative relationship, which was consistent with what was found by Nath, Cannon, and Philipp (2019) for both muscles and Sato et al. (2020) for the *corrugator*. Activity of both muscles (*corrugator* and *levator*) could predict select EsSense 25 terms, having a positive relationship with negative valence terms, and the reverse for positive valence terms. In addition, the *zygomaticus* was unable to predict hedonic liking for both product sets, consistent with Sato et al. (2020), but did have the ability to predict certain EsSense 25 terms. It is possible that with more participants who were emotionally expressive, there would be correlations between the implicit and explicit measures, as with the present study there may be no overlap between the participants that used some terms and those who reacted with facial expressions. Conversely, it is possible that the lack of correlations indicates that the implicit measures of emotion are recording something different to the explicit measures, possibly the subconscious aspects of emotion (Berridge 2018). The implication of this is that the inclusion of implicit measures of emotion could allow industry greater insights into what drives choice for groups of consumers that are facially expressive.

4.4 Do implicit and explicit measures of emotional response and liking differentiate milk and yoghurt differently?

There is evidence in the literature to suggest that implicit measures of emotional response such as FEA and EDA are better able to differentiate between stimuli with a wide range of sensory characteristics than those that are more similar Kaneko et al. (2019). The liking and facial EMG results support this, with greater differentiation seen within the yoghurt products than the milk. Overall, more yoghurt products were discriminated by liking, and all three facial EMG muscles were able to differentiate yoghurt compared to two for milk. This is likely because of the larger differences in sensory characteristics of the yoghurt compared to the milk products and is consistent with the findings of Kaneko et al. (2019) where more methods of measuring emotional response were able to differentiate diluted vinegar from a range of regular drinks than between the regular drinks. Whilst it is possible that there is truly no difference in consumer response between similar products such as the milks used in this study, it is more likely that none of the measures of liking and emotion were sensitive enough to detect these differences alone. Combining multiple implicit and explicit measures has been shown to be able to predict liking and consumer preference for vegetable juices better than the individual measures (Samant and Seo 2019). This approach could be applied to products with small differences in sensory characteristics and may improve product differentiation and prediction of liking and consumer choice.

4.5 Are the selected implicit and explicit measures impacted by the use of an individually composed written evoked scenario?

Changing the consumption context has been found to affect consumers' emotional response (Dorado, Chaya, et al. 2016, Low et al. 2021, Xu et al. 2019). However, this was not seen in the current study, which found that only the EsSense 25 term 'Good' had a significant effect of scenario for the milk products. This could be due to the unfamiliar nature of having electrodes attached, which may cause a greater difference between the first and second sessions rather than the presence or absence of the written scenario. Low emotional engagement with the products may be the cause, however the participants were less engaged with milks, but the only significant effect of scenario was found in this product set. Alternatively, changing the consumption context through an evoked scenario written by the individual participant may have been the reason that there was no effect. De Wijk et al. (2019a) suggested that similarities between consumption contexts may have more impact than the differences. In the present study, participants followed the same procedure, used the same computer and mouse and tasted samples of the same size across both consumption contexts, with the only difference being the additional task of writing the scenario. This highlights a limitation of written scenarios, as it relies on a participant's imagination and engagement with writing a scenario that is relevant to them. Several methods that do not rely on participants' imagination have been used in other studies including using images or descriptions of situations (Piqueras-Fiszman and Jaeger 2014a), a change in physical location (Xu et al. 2019) and the use of virtual reality (VR) to change the location (Low et al. 2021). For this study, using VR was not possible because the headsets obscure the face and would likely rest on the EMG electrodes, and changing the location was not possible due to equipment constraints. This shows that more research is needed for how to evoke a scenario that is engaging but not disruptive to other measures.

4.6 Suitability of the procedures for the measures used

Whilst there were no significant problems with the data collection for the facial EMG or explicit measures, there were some issues with timing of product consumption with participants tasting samples early or late. Leitch et al. (2015) screened participants for their ability to follow instructions in a practice session before data collection, which is an approach that may be beneficial when precise timings of consumption are important to the processing of the data. For EDA, it is difficult to determine whether the lack of significant product main effects is due to collection errors or whether there was no real difference between products. It is possible that there was too much noise from hand movement in the EDA, and that asking participants to keep the hand with the sensors stationary may yield cleaner data and better results (Rita, Guerreiro, and Omarji 2020).

There were many difficulties faced in the recording of facial emotions using facial expression analysis in this study. One of which was that the lighting in the laboratory was not ideal and cast shadows onto the face, causing the image to be too dark for the software to interpret which is consistent with what was recorded in De Wijk et al. (2019a). Other factors appeared to affect the software's ability to recognise facial emotions such as skin colour and facial features. Leitch et al. (2015) screened potential participants to remove those with glasses, beards, and other facial features that might interfere with the ability of the facial expression analysis software to recognise facial expressions. This was not done in this investigation, and in addition to participants with those features every participant had eight electrodes on their face which further impeded the software's ability to identify a face and record facial emotions. Another common problem with recording facial emotions using video is participants obscuring their face from the camera holding the cup in front of their face before or after tasting the sample, (Zhi, Cao, and Cao 2017, Samant and Seo 2019). In studies without EMG, facial expression analysis software can yield emotional profiles that can differentiate products (de Wijk et al. 2019b, Juodeikiene et al. 2018), however some data is generally lost due to video quality. When used as an explicit measure of emotion, facial expression analysis is highly correlated to facial EMG, indicating its potential as a less invasive alternative, however more investigation is needed to determine whether this is a viable alternative.

4.7 Limitations of the investigation due to the COVID-19 pandemic

Restrictions in face-to-face contact due to COVID-19 meant that there was limited time available to run a full pilot experiment before data collection, so this was done with only two participants. Because of this short pilot, some issues were discovered during data collection, and changes were made during this time. Further, there were few participants who were not employed by or studying at Massey, possibly due to concerns about social distancing during the study. In addition, the time frame to collect data within was condensed with only nine weeks due to the delays and limited time before closure of the Massey University campus for the holidays in mid-December 2020.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Hedonic liking was more discriminating for milk and yoghurt products than explicit and implicit emotional response. Nevertheless, including emotional response alongside liking gives a different perspective to how participants respond to products. Self-report emotional response was the most discriminating of the emotion measures, however all these measures were hindered by participants' lack of emotional engagement with the products. Whilst some EsSense 25 terms were able to differentiate between products, others were used by less than half of participants, demonstrating low emotional engagement with the lexicon. Facial EMG was the most discriminating of the implicit methods, with the *corrugator* and *levator* muscles differentiating both milk and yoghurt products, and the *zygomaticus* differentiating yoghurts. However, the ability of the muscles to differentiate products was dependent on participant, as each muscle had a cluster of participants with little difference in mean activity across all products. Despite this, facial EMG has potential to be a useful measure of emotional response, as the corrugator and *levator* muscles were found to have the ability to predict liking and some EsSense 25 lexicon terms. The other implicit measures used in this study were less promising, with the EDA found to be a poor measure of emotional response to milk and yoghurt, and no useful data able to be extracted from the FEA software. More investigation is needed to assess the suitability of FEA software in measuring emotional response to tasted samples.

Despite the small differences in sensory characteristics, differentiation was seen between some milk products for the liking, self-report emotional response, and *corrugator* and *levator* muscles. However, greater differentiation was seen between the yoghurt products for the self-report emotional response and liking, and all three facial EMG muscles could differentiate the yoghurt.

The use of an individually composed written scenario did not have a significant effect on the liking or emotional response except for the self-report term 'Good' for the milk products. This could be due to a lack of engagement with the products or a limitation of the method of evoking the context, or both.

5.2 RECOMMENDATIONS

This thesis does not fully explore the data collected during this project; further insights could be gained from the following:

- Further data analysis on the data set using multivariate techniques.
- Investigating the ability of clusters of participants for hedonic liking and emotional response to differentiate products.

- Comparing the liking clusters to participants' most consumed milk and yoghurt products from responses in the recruitment questionnaire.

For the use of these methods for measuring emotional response in future studies, the following is recommended:

- A product specific lexicon should be created and used for the explicit measurement of emotional response, particularly for products that have low emotional engagement like milk and yoghurt.
- Where precise timing is important (for implicit measures), have participants follow along with a video or avatar instead of relying on written instructions to maintain better timing. Alternatively, screen potential participants for ability to follow instructions.
- Further investigate FEA software to determine the strengths and limitations of this method before use in measuring emotional response to tasted samples. Researchers should conduct extensive piloting with a variety of participants to ensure that the tasting procedures and other methodology are compatible with the FEA software.
- If using facial EMG or FEA, screen potential participants for emotional expressiveness, or collect data from a larger number of participants.
- For products with small differences in sensory characteristics, responses should be collected from a larger number of participants.
- More engaging methods of evoking scenarios to change the consumptions context should be investigated.

REFERENCES

Ahn, Hyung-il, and Rosalind W. Picard. 2014. "Measuring Affective-Cognitive Experience and Predicting Market Success." *Ieee Transactions on Affective Computing* 5 (2):173-186. doi: 10.1109/taffc.2014.2330614.

Alvarez-Pato, V. M., C. N. Sanchez, J. Dominguez-Soberanes, D. E. Mendoza-Perez, and R. Velazquez. 2020. "A multisensor data fusion approach for predicting consumer acceptance of food products." *Foods* 9 (6):774-774. doi: 10.3390/foods9060774.

Ares, Gastón, Fernanda Bruzzone, Leticia Vidal, Rafael Silva Cadena, Ana Giménez, Benedicte Pineau, Denise C. Hunter, Amy G. Paisley, and Sara R. Jaeger. 2014. "Evaluation of a rating-based variant of check-all-that-apply questions: Rate-all-that-apply (RATA)." *Food Quality and Preference* 36:87-95. doi: https://doi.org/10.1016/j.foodqual.2014.03.006.

Barrett, Lisa Feldman. 2016. "The theory of constructed emotion: an active inference account of interoception and categorization." *Social Cognitive and Affective Neuroscience* 12 (1):1-23. doi: 10.1093/scan/nsw154.

Barrett, Lisa Feldman, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements." *Psychological Science in the Public Interest* 20 (1):1-68. doi: 10.1177/1529100619832930.

Benedek, M., and C. Kaernbach. 2010. "Decomposition of skin conductance data by means of nonnegative deconvolution." *Psychophysiology* 47 (4):647-658. doi: 10.1111/j.1469-8986.2009.00972.x.

Berridge, Kent C. 2018. "Evolving Concepts of Emotion and Motivation." *Frontiers in Psychology* 9 (1647). doi: 10.3389/fpsyg.2018.01647.

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

Bhumiratana, Natnicha, Koushik Adhikari, and Edgar Chambers. 2014. "The development of an emotion lexicon for the coffee drinking experience." *Food Research International* 61:83-92. doi: https://doi.org/10.1016/j.foodres.2014.03.008.

Brouwer, A. M., M. A. Hogervorst, M. Grootjen, J. B. F. van Erp, and E. H. Zandstra. 2017. "Neurophysiological responses during cooking food associated with different emotions." *Food Quality and Preference* 62:307-316. doi: 10.1016/j.foodqual.2017.03.005.

Brouwer, A. M., M. A. Hogervorst, J. B. F. van Erp, M. Grootjen, E. van Dam, and E. H. Zandstra. 2019. "Measuring cooking experience implicitly and explicitly: Physiology, facial expression and subjective ratings." *Food Quality and Preference* 78. doi: 10.1016/j.foodqual.2019.103726.

Chen, Y., Q. Gao, Q. Lv, N. Qian, and L. Ma. 2018. "Comparing measurements for emotion evoked by oral care products." *International Journal of Industrial Ergonomics* 66:119-129. doi: 10.1016/j.ergon.2018.02.013.

Coppin, Géraldine, and David Sander. 2016. "Theoretical Approaches to Emotion and Its Measurement." In *Emotion Measurement*, edited by Herbert L. Meiselman, 3-30. Woodhead Publishing.

Crist, C. A., S. E. Duncan, E. A. Arnade, K. A. Leitch, S. F. O'Keefe, and D. L. Gallagher. 2018. "Automated facial expression analysis for emotional responsivity using an aqueous bitter model." *Food Quality and Preference* 68:349-359. doi: https://doi.org/10.1016/j.foodqual.2018.04.004.

Danner, Lukas, Liudmila Sidorkina, Max Joechl, and Klaus Duerrschmid. 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: 10.1016/j.foodqual.2013.01.004.

Darwin, Charles. 1872. *The expression of the emotions in man and animals by Charles Darwin*: Murray.

De Beuckelaer, Alain, Machiel Zeeman, and Hans Van Trijp. 2015. "Assessment of the cross-national validity of an End-anchored 9-point hedonic product liking scale." *Quality & Quantity* 49 (3):1267-1286. doi: 10.1007/s11135-014-0049-0.

De Wijk, R. A., W. He, M. G. J. Mensink, R. H. G. Verhoeven, and C. De Graaf. 2014. "ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks." *PLoS ONE* 9 (4). doi: 10.1371/journal.pone.0093823.

de Wijk, R., A., D. Kaneko, G. B. Dijksterhuis, M. van Zoggel, I. Schiona, M. Visalli, and E. H. Zandstra. 2019a. "Food perception and emotion measured over time in-lab and in-home." *Food Quality and Preference* 75:170-178. doi: https://doi.org/10.1016/j.foodqual.2019.02.019.

De Wijk, R. A., D. Kaneko, G. B. Dijksterhuis, M. van Zoggel, I. Schiona, M. Visalli, and E. H. Zandstra. 2019b. "Food perception and emotion measured over time in-lab and in-home." *Food Quality and Preference* 75:170-178. doi: 10.1016/j.foodqual.2019.02.019.

Dimberg, U. 1990. "Facial electromyography and emotional reactions." *Psychophysiology* 27 (5):481-94. doi: 10.1111/j.1469-8986.1990.tb01962.x.

Dolensek, Nejc, Daniel A. Gehrlach, Alexandra S. Klein, and Nadine Gogolla. 2020. "Facial expressions of emotion states and their neuronal correlates in mice." *Science* 368 (6486):89-+. doi: 10.1126/science.aaz9468.

Donato, G., M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. 1999. "Classifying facial actions." *Ieee Transactions on Pattern Analysis and Machine Intelligence* 21 (10):974-989. doi: 10.1109/34.799905.

Dorado, R., C. Chaya, A. Tarrega, and J. Hort. 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., C. Pérez-Hugalde, A. Picard, and C. Chaya. 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.

Duchenne de Boulogne, Guillaume-Benjamin. 1862. The Mechanism of Human Facial Expression, ed. and trans. RA Cuthbertson. Cambridge: Cambridge University Press. doi.

Ekman, P, WV Friesen, and J Hager. 1978. "Facial action coding system: A technique for the measurement of facial movement."

Garcia-Burgos, D., and M. C. Zamora. 2013. "Facial affective reactions to bitter-tasting foods and body mass index in adults." *Appetite* 71:178-186. doi: https://doi.org/10.1016/j.appet.2013.08.013.

Gonzalez Viejo, Claudia, Sigfredo Fuentes, Kate Howell, Damir D. Torrico, and Frank R. Dunshea. 2019. "Integration of non-invasive biometrics with sensory analysis techniques to assess acceptability of beer by consumers." *Physiology & Behavior* 200:139-147. doi: https://doi.org/10.1016/j.physbeh.2018.02.051.

Gunaratne, Thejani M., Sigfredo Fuentes, Nadeesha M. Gunaratne, Damir D. Torrico, Claudia Gonzalez Viejo, and Frank R. Dunshea. 2019. "Physiological Responses to Basic Tastes for Sensory Evaluation of Chocolate Using Biometric Techniques." *Foods* 8 (7). doi: 10.3390/foods8070243.

Gutjar, Swetlana, Jelle R. Dalenberg, Cees de Graaf, René A. de Wijk, Aikaterini Palascha, Remco J. Renken, and Gerry Jager. 2015. "What reported food-evoked emotions may add: A model to predict consumer food choice." *Food Quality and Preference* 45:140-148. doi: https://doi.org/10.1016/j.foodqual.2015.06.008.

Jaeger, S. R., D. Jin, D. C. Hunter, C. M. Roigard, and D. I. Hedderley. 2020. "Multi-response approaches in product-focused investigations: Methodological variations across three case studies." *Food Research International* 132. doi: 10.1016/j.foodres.2020.109113.

Jaeger, Sara R., S. M. Lee, K. O. Kim, S. L. Chheang, D. Jin, and G. Ares. 2017. "Measurement of product emotions using emoji surveys: Case studies with tasted foods and beverages." *Food Quality and Preference* 62:46-59. doi: 10.1016/j.foodqual.2017.05.016.

Jaeger, Sara R., Marianne Swaney-Stueve, Sok L. Chheang, Denise C. Hunter, Benedicte Pineau, and Gastón Ares. 2018. "An assessment of the CATA-variant of the EsSense Profile®." *Food Quality and Preference* 68:360-370. doi: https://doi.org/10.1016/j.foodqual.2018.04.005.

Juodeikiene, Grazina, Daiva Zadeike, Dovile Klupsaite, Darius Cernauskas, Elena Bartkiene, Vita Lele, Vesta Steibliene, and Virginija Adomaitiene. 2018. "Effects of emotional responses to certain foods on the prediction of consumer acceptance." *Food Research International* 112:361-368. doi: https://doi.org/10.1016/j.foodres.2018.06.064.

Kaneko, D., M. Hogervorst, A. Toet, J. B. F. van Erp, V. Kallen, and A. M. Brouwer. 2019. "Explicit and implicit responses to tasting drinks associated with different tasting experiences." *Sensors (Switzerland)* 19 (20). doi: 10.3390/s19204397.

Kanjanakorn, Alisa, and Jeehyun Lee. 2017. "Examining emotions and comparing the EsSense Profile® and the Coffee Drinking Experience in coffee drinkers in the natural environment." *Food Quality and Preference* 56:69-79. doi: https://doi.org/10.1016/j.foodqual.2016.09.014.

Kenney, E., and K. Adhikari. 2016. "Recent developments in identifying and quantifying emotions during food consumption." *Journal of the science of food and agriculture* 96 (11):3627-3630. doi: 10.1002/jsfa.7717.

Kim, Seon-Ho, and Jae-Hee Hong. 2020. "The effects of contexts on consumer emotions and acceptance of a domestic food and an unfamiliar ethnic food: a cross-cultural comparison between Chinese and Korean consumers." *Food Science and Biotechnology* 29 (12):1705-1718. doi: 10.1007/s10068-020-00827-2.

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

King, Silvia C., Herbert L. Meiselman, Annette W. Hottenstein, Therese M. Work, and Valerie Cronk. 2007. "The effects of contextual variables on food acceptability: A confirmatory study." *Food Quality and Preference* 18 (1):58-65. doi: https://doi.org/10.1016/j.foodqual.2005.07.014.

Kong, Yanzhuo, Chetan Sharma, Madhuri Kanala, Mishika Thakur, Lu Li, Dayao Xu, Roland Harrison, and Damir D. Torrico. 2020. "Virtual Reality and Immersive Environments on Sensory Perception of Chocolate Products: A Preliminary Study." *Foods* 9 (4). doi: 10.3390/foods9040515.

Köster, Egon P., and Jozina Mojet. 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: https://doi.org/10.1016/j.foodres.2015.04.006.

Kostyra, Eliza, Michał Rambuszek, Bożena Waszkiewicz-Robak, Wacław Laskowski, Tadeusz Blicharski, and Ewa Poławska. 2016. "Consumer facial expression in relation to smoked ham with the use of face reading technology. The methodological aspects and informative value of research results." *Meat Science* 119:22-31. doi: https://doi.org/10.1016/j.meatsci.2016.04.018.

Kostyra, Eliza, Grayna Wasiak-Zys, Michal Rambuszek, and Bozena Waszkiewicz-Robak. 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: 10.1016/j.lwt.2015.08.008.

Kring, A. M., and D. M. Sloan. 2007. "The Facial Expression Coding System (FACES): Development, validation, and utility." *Psychological Assessment* 19 (2):210-224. doi: 10.1037/1040-3590.19.2.210.

Larosa, C. P., C. F. Balthazar, J. T. Guimarâes, R. S. Rocha, R. Silva, T. C. Pimentel, D. Granato, M. C. K. H. Duarte, M. C. Silva, M. Q. Freitas, A. G. Cruz, and E. A. Esmerino. 2021. "Sheep milk kefir sweetened with different sugars: Sensory acceptance and consumer emotion profiling." *Journal of Dairy Science* 104 (1):295-300. doi: 10.3168/jds.2020-18702.

Le Goff, G., and J. Delarue. 2017. "Non-verbal evaluation of acceptance of insect-based products using a simple and holistic analysis of facial expressions." *Food Quality and Preference* 56:285-293. doi: 10.1016/j.foodqual.2016.01.008.

Leitch, K. A., S. E. Duncan, S. O'Keefe, R. Rudd, and D. L. Gallagher. 2015. "Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis." *Food Research International* 76:283-292. doi: 10.1016/j.foodres.2015.04.039.

Low, J. Y. Q., V. H. F. Lin, L. Jun Yeon, and J. Hort. 2021. "Considering the application of a mixed reality context and consumer segmentation when evaluating emotional response to tea break snacks." *Food Quality and Preference* 88. doi: 10.1016/j.foodqual.2020.104113.

Martinez, B., M. F. Valstar, B. H. Jiang, and M. Pantic. 2019. "Automatic Analysis of Facial Actions: A Survey." *Ieee Transactions on Affective Computing* 10 (3):325-347. doi: 10.1109/taffc.2017.2731763.

Mehta, Annu, Chetan Sharma, Madhuri Kanala, Mishika Thakur, Roland Harrison, and Damir Dennis Torrico. 2021. "Self-Reported Emotions and Facial Expressions on Consumer Acceptability: A Study Using Energy Drinks." *Foods* 10 (2):330.

Mojet, Jozina, Klaus Dürrschmid, Lukas Danner, Max Jöchl, Raija-Liisa Heiniö, Nancy Holthuysen, and Egon Köster. 2015. "Are implicit emotion measurements evoked by food unrelated to liking?" *Food Research International* 76:224-232. doi: https://doi.org/10.1016/j.foodres.2015.06.031.

Mora, M., E. Urdaneta, and C. Chaya. 2018. "Emotional response to wine: Sensory properties, age and gender as drivers of consumers' preferences." *Food Quality and Preference* 66:19-28. doi: https://doi.org/10.1016/j.foodqual.2017.12.015.

Mora, María, Elena Urdaneta, and Carolina Chaya. 2019. "Effect of personality on the emotional response elicited by wines." *Food Quality and Preference* 76:39-46. doi: https://doi.org/10.1016/j.foodqual.2019.03.015.

Nath, E. C., P. R. Cannon, and M. C. Philipp. 2019. "An unfamiliar social presence reduces facial disgust responses to food stimuli." *Food Research International* 126. doi: 10.1016/j.foodres.2019.108662.

Nath, E. C., P. R. Cannon, and M. C. Philipp. 2020. "Co-acting strangers but not friends influence subjective liking and facial affective responses to food stimuli." *Food Quality and Preference* 82. doi: 10.1016/j.foodqual.2019.103865.

Nestrud, Michael A., Herbert L. Meiselman, Silvia C. King, Larry L. Lesher, and Armand V. Cardello. 2016. "Development of EsSense25, a shorter version of the EsSense Profile®." *Food Quality and Preference* 48:107-117. doi: https://doi.org/10.1016/j.foodqual.2015.08.005.

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

Nguyen, Ha, and Wendy V. Wismer. 2019. "A comparison of sensory attribute profiles and liking between regular and sodium-reduced food products." *Food Research International* 123:631-641. doi: https://doi.org/10.1016/j.foodres.2019.05.037.

Oliveira, D., H. De Steur, S. Lagast, X. Gellynck, and J. J. Schouteten. 2020. "The impact of calorie and physical activity labelling on consumer's emo-sensory perceptions and food choices." *Food Research International* 133. doi: 10.1016/j.foodres.2020.109166.

Piqueras-Fiszman, B., and S. R. Jaeger. 2014a. "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, Betina, and Sara R. Jaeger. 2014b. "The impact of evoked consumption contexts and appropriateness on emotion responses." *Food Quality and Preference* 32:277-288. doi: https://doi.org/10.1016/j.foodqual.2013.09.002.

Richins, M. L. 1997. "Measuring emotions in the consumption experience." *Journal of Consumer Research* 24 (2):127-146. doi: 10.1086/209499.

Rita, Paulo, Joao Guerreiro, and Muhamad Omarji. 2020. "Autonomic emotional responses to food: Private label brands versus National Brands." *Journal of Consumer Behaviour*. doi: 10.1002/cb.1874.

Rocha, C., R. C. Lima, A. P. Moura, T. Costa, and L. M. Cunha. 2019. "Implicit evaluation of the emotional response to premium organic herbal infusions through a temporal dominance approach: Development of the temporal dominance of facial emotions (TDFE)." *Food Quality and Preference* 76:71-80. doi: 10.1016/j.foodqual.2019.04.001.

Rohlfing, B. 1997. Alaska's wild Denali [Motion picture]. Anchorage, AK: Alaska Video Postcards. Inc.

Ruth, Julie, Frederic Brunel, and Cele Otnes. 2002. *Linking Thoughts to Feelings: Investigating Cognitive Appraisals and Consumption Emotions in a Mixed-Emotions Context*. Vol. 30.

Samant, S. S., M. J. Chapko, and H. S. Seo. 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: 10.1016/j.foodres.2017.07.021.

Samant, S. S., and H. S. Seo. 2019. "Using both emotional responses and sensory attribute intensities to predict consumer liking and preference toward vegetable juice products." *Food Quality and Preference* 73:75-85. doi: 10.1016/j.foodqual.2018.12.006.

Samant, Shilpa S., and Han-Seok Seo. 2020. "Influences of sensory attribute intensity, emotional responses, and non-sensory factors on purchase intent toward mixed-vegetable juice products under informed tasting condition." *Food Research International* 132. doi: 10.1016/j.foodres.2020.109095.

Sander, David. 2013. *The Cambridge handbook of human affective neuroscience.*, 5-53. New York, NY, US: Cambridge University Press.

Sargent, Amanda, Jan Watson, Hongjun Ye, Rajneesh Suri, and Hasan Ayaz. 2020. "Neuroergonomic Assessment of Hot Beverage Preparation and Consumption: An EEG and EDA Study." *Frontiers in Human Neuroscience* 14. doi: 10.3389/fnhum.2020.00175.

Sato, W., K. Minemoto, A. Ikegami, M. Nakauma, T. Funami, and T. Fushiki. 2020. "Facial EMG correlates of subjective hedonic responses during food consumption." *Nutrients* 12 (4). doi: 10.3390/nu12041174.

Sato, Wataru, Sakiko Yoshikawa, and Tohru Fushiki. 2020. "Facial EMG Activity Is Associated with Hedonic Experiences but not Nutritional Values While Viewing Food Images." *Nutrients* 13 (1). doi: 10.3390/nu13010011.

Schilling, M. W., and P. C. Coggins. 2007. "Utilization of agglomerative hierarchical clustering in the analysis of hedonic scaled consumer acceptability data." *Journal of Sensory Studies* 22 (4):477-491. doi: 10.1111/j.1745-459X.2007.00121.x.

Schouteten, J., J., H. De Steur, S. De Pelsmaeker, S. Lagast, I. De Bourdeaudhuij, and X. Gellynck. 2015. "An integrated method for the emotional conceptualization and sensory characterization of food products: The EmoSensory® Wheel." *Food Research International* 78:96-107. doi: https://doi.org/10.1016/j.foodres.2015.11.001.

Singh, A., and H. S. Seo. 2020. "Sample temperatures can modulate both emotional responses to and sensory attributes of tomato soup samples." *Food Quality and Preference* 86. doi: 10.1016/j.foodqual.2020.104005.

Soussignan, R., B. Schaal, V. Boulanger, S. Garcia, and T. Jiang. 2015. "Emotional communication in the context of joint attention for food stimuli: Effects on attentional and affective processing." *Biological Psychology* 104:173-183. doi: 10.1016/j.biopsycho.2014.12.006.

Spinelli, S., C. Masi, C. Dinnella, G. Zoboli, P., and E. Monteleone. 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: https://doi.org/10.1016/j.foodqual.2013.11.009.

Torrico, D. D., S. Fuentes, C. Gonzalez Viejo, H. Ashman, N. M. Gunaratne, T. M. Gunaratne, and F. R. Dunshea. 2018. "Images and chocolate stimuli affect physiological and affective responses of consumers: A cross-cultural study." *Food Quality and Preference* 65:60-71. doi: 10.1016/j.foodqual.2017.11.010.

van Bommel, R., M. Stieger, M. Visalli, R. de Wijk, and G. Jager. 2020. "Does the face show what the mind tells? A comparison between dynamic emotions obtained from facial expressions and Temporal Dominance of Emotions (TDE)." *Food Quality and Preference* 85. doi: 10.1016/j.foodqual.2020.103976.

Watson, David. 1988. "The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect." *Journal of Personality and Social Psychology* 55 (1):128-141. doi: 10.1037/0022-3514.55.1.128.

Williams, EJ. 1949. "Experimental Designs Balanced for the Estimation of Residual Effects of Treatments." *Australian Journal of Chemistry* 2 (2):149-168. doi: https://doi.org/10.1071/CH9490149.

Worch, Thierry, Fiorella Sinesio, Elisabetta Moneta, Silvia Abba, Lise Dreyfuss, Jean A. McEwan, and Christelle Porcherot-Lassallette. 2020. "Influence of different test conditions on the emotional responses elicited by beers." *Food Quality and Preference* 83. doi: 10.1016/j.foodqual.2020.103895.

Xu, Yun, Nazimah Hamid, Daniel Shepherd, Kevin Kantono, and Charles Spence. 2019. "Changes in flavour, emotion, and electrophysiological measurements when consuming chocolate ice cream in different eating environments." *Food Quality and Preference* 77:191-205. doi: 10.1016/j.foodqual.2019.05.002.

Zhi, R., L. Cao, and G. Cao. 2017. "Asians' Facial Responsiveness to Basic Tastes by Automated Facial Expression Analysis System." *Journal of Food Science* 82 (3):794-806. doi: 10.1111/1750-3841.13611.

Zhi, R., J. Wan, D. Zhang, and W. Li. 2018. "Correlation between hedonic liking and facial expression measurement using dynamic affective response representation." *Food Research International* 108:237-245. doi: 10.1016/j.foodres.2018.03.042.

Zokaityte, Egle, Vita Lele, Vytaute Starkute, Paulina Zavistanaviciute, Darius Cernauskas, Dovile Klupsaite, Modestas Ruzauskas, Juste Alisauskaite, Alma Baltrusaityte, Mantvydas Dapsas, Karolina Siriakovaite, Simonas Trunce, Raquel P. F. Guine, Pranas Viskelis, Vesta Steibliene, and Elena Bartkiene. 2020. "Antimicrobial, Antioxidant, Sensory Properties, and Emotions Induced for the Consumers of Nutraceutical Beverages Developed from Technological Functionalised Food Industry By-Products." *Foods* 9 (11). doi: 10.3390/foods9111620.

Appendix 1 – Participant recruitment

Statement of ethics approval

This project has been reviewed and approved by the Massey University Human Ethics Committee: Southern A, Application 20/30. If you have any concerns about the conduct of this research, please contact Dr Negar Partow, Chair, Massey University Human Ethics Committee: Southern A, telephone 04 801 5799 x 63363, email humanethicsoutha@massey.ac.nz

Recruitment poster

Figure A1: Poster used to recruit participants for this study.

APPENDIX 2 – INFORMATION SHEET

Evaluating emotional response to foods

INFORMATION SHEET

Researcher(s) Introduction: This activity is run by a Master's student Rachel Taylor and is under supervision of Professor Joanne Hort, Dr Peter Cannon, and Dr Amanda Dupas de Matos.

Project Description and Invitation: This study invites you to participate in a sensory activity to help investigate emotional response to foods. You will be offered a goodie bag and a $50 disturbance allowance as a thank you for your time on completion of the full study. The purpose of this project will be explained at the end of the session.

Participant Identification and Recruitment: Potential participants will be contacted through the Feast consumer database or will have responded to a recruitment poster. Anyone may participate who is 18 years old or older, who does not have any allergies or intolerances to the foods listed at the end of this document, and meets the following criteria:

- Not pregnant or lactating
- Not allergic or intolerant to the following: Cow's milk, live yoghurt cultures, cream, gelatin, modified corn starch, sugar, crackers (gluten free available).

Project Procedures:

For this project, you will be required to attend two sessions, each expected to take about 60-75 minutes. During the sessions you will be tasting small portions (10g) of different milk and plain yoghurt samples (equating to not more than 400kJ in total) with sensors attached to your face and fingers using stickers similar to an adhesive plaster and following good hygienic practice. You will rate your liking and answer an emotion questionnaire after tasting each sample. Prior to the application of the sensors, cleaning and gentle exfoliation of the face will be required, this will be done by the researcher. In rare cases there may be a reaction to the facial scrubbing in the cleaning steps in preparation for the electrode application. If this occurs, the experiment will be stopped, and you will be recommended to see your GP if the reaction causes concern.

As part of the session there will be a video to watch that includes imagery of heights. At the end of the session, you will be instructed to make various facial expressions before the sensors are removed. Video will also be recorded during the test session. The video will be used as a further mechanism to assess emotional response and will be analyzed using facial expression analysis software by the researcher. Please note we require participants to keep the details of the session confidential.

Data Management: Contact details of participants and the research data/videos will be stored in separate password protected files on a Massey password protected server, only accessible by the research team members. Collected data (but no personal information) will be used in a confidential report and shared with Fonterra who provided funding for the project. Confidentiality of identity will be preserved in all such publications. Your data will be kept for 7 years before being destroyed.

Participant's Rights: You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- decline to answer the question;
- withdraw from the activity;
- ask any questions about the activity at any time during participation;
- provide information on the understanding that your name will not be used unless you give permission to the researcher.

Project Contacts: Please contact the researcher(s) and/or supervisor(s) if you have any questions about the project:

- *Master student: Rachel Taylor*
- *Supervisor: Joanne Hort (J.hort@massey@ac.nz)*

Ethics Approval

This project has been reviewed and approved by the Massey University Human Ethics Committee: Southern A, Application 20/30. If you have any concerns about the conduct of this research, please contact Dr Negar Partow, Chair, Massey University Human Ethics Committee: Southern A, telephone 04 801 5799 x 63363, email humanethicsoutha@massey.ac.nz

Compensation for Injury

If physical injury results from your participation in this study, you should visit a treatment provider to make a claim to ACC as soon as possible. ACC cover and entitlements are not automatic, and your claim will be assessed by ACC in accordance with the Accident Compensation Act 2001. If your claim is accepted, ACC must inform you of your entitlements, and must help you access those entitlements. Entitlements may include, but not be limited to, treatment costs, travel costs for rehabilitation, loss of earnings, and/or lump sum for permanent impairment. Compensation for mental trauma may also be included, but only if this is incurred as a result of physical injury.

If your ACC claim is not accepted, you should immediately contact the researcher. The researcher will initiate processes to ensure you receive compensation equivalent to that to which you would have been entitled had ACC accepted your claim.

APPENDIX 3 – SCRIPTS USED IN CONDUCTING EXPERIMENTAL SESSIONS

Script 1: Meeting participant in waiting area

Hi (participant's name), I'm (researcher name). I'm running this study. Thank you for volunteering. Please follow me to the laboratory.

Once in lab

You can put your jacket/bag/etc over here (show)
Have you had a chance to read the information sheet?

 if yes: do you have any questions?

 if no: I have a copy of it here. Please read it and let me know if you have any questions.

After any questions/if they don't have any questions: If you're happy to do the experiment still and agree with the conditions please sign the consent form and we'll get started.

The first thing we need to do is get you to wash your face using this cleanser (hand them Cetaphil bottle and cotton pads). This is to get rid of any residue moisturiser, makeup, skin oil that might affect the electrode signals. Please, wash your whole face but pay particular attention to these areas (gesture to forehead, just above left eyebrow, the left side of the nose, and the middle of the cheek). Please also wash your hands with soap and water before you leave the bathroom and come back to the lab with the cleanser bottle once you're done.
Take them to the bathroom

Script 2: Once they are back in the lab

Please sit in this chair over here (show chair in front of participant desk)
Before I start putting the electrodes on, I'll put this headband on you first. This is what the electrodes are plugged into and this sends a Bluetooth signal to the machine that records the muscle activity (put headband on). Does this feel comfortable? It's not too tight? Does it feel like it will fall off if you move your head? (adjust if necessary)
I'm going to start with the cleaning steps that we need to do before I can apply the electrodes, please let me know if anything is uncomfortable or painful.
I'm going to use an alcohol wipe on the sites where electrodes will go to remove any moisturiser, makeup, skin oil that might still be on the skin. (use alcohol wipes, ask participant to close their eyes before wiping the *levator* area)
Now I am going to gently exfoliate the areas where the electrodes will be applied, using this (show abrasive pad). This is to remove some of the dead skin cells that make up the outer layers of skin so that the electrodes can get a better signal (exfoliate).
I'm going to use an alcohol wipe again to wipe off the dead skin cells that we removed in the exfoliating step. This might sting, but hopefully it won't (do so).
I'm going to start putting the electrodes on. First, I'm going to rub a little bit of the electrode gel onto the areas so that the skin can absorb it and not absorb all the gel from the electrodes when I put them on. Then, I'll stick the electrodes on and check that the signal is good.

 If electrodes have a bad signal: I'll need to re-apply this pair because the signal isn't where it needs to be. It looks like there might be a bubble in the gel or something interfering with the signal. I'll just re-clean the area briefly to make sure it works this next time.

EDA electrode application:

There's one more set of electrodes to be applied, but these ones go on your hand. Are you right or left-handed?
Like the headband, there's an armband that we need to put on the same arm as the electrodes (put on). How does this feel? Does it feel secure? (adjust if necessary)
The cleaning steps for these electrodes are a lot simpler, it only needs an alcohol wipe and then some gel and tape to hold it in place (do these then put the electrodes on).

Script 3: starting the experiment

I'll explain the task to you, and then I'll go get the samples that you'll need for it.
What you will be doing is following instructions on the screen. First, you'll be asked to watch a video and then answer a questionnaire about how it made you feel. (If scenario, otherwise SKIP this paragraph)

Then you will be asked to describe a time of day and situation when you would generally drink milk and type it into the box on screen. Later in the experiment you will be asked to describe a scenario specific to yoghurt as well. There is no right or wrong answer, we want you to imagine a scenario that is applicable to you and give us as much detail as you can in five minutes.
For example: if we were asking about ice cream, I might answer "Time of day: evening" and "situation: It's 8pm and I feel like I deserve a treat because it's been a stressful day. I rummage in the kitchen and there isn't any chocolate, so I grab some ice cream from the freezer."
There will be another example on the screen as well.

Sample procedure

After this you will be instructed to pick up a specific sample, please pick up the sample with the hand you don't have electrodes on. You will then be asked to taste the sample by drinking one big sip (or one big spoonful for the yoghurt samples). Then, you will be asked to answer a questionnaire about how it made you feel and how much you liked it. The samples should be in the right order on the tray when I bring them in, but check the codes match before you taste it, just in case. There will be a slight delay between samples, and once the milk samples are done, there will be a short break where I will bring in the yoghurt samples.

Before the end of the experiment there will be another video to watch and then I'll come back in to take you through the last part of the task.

If you have any issues, call out to me. I'll be through the door there and will come sort it out. Any questions?

Script 4: Maximum voluntary contractions

This last part will be pretty quick. You'll need to contract your face in various ways as much as you can. I'll demonstrate a facial expression that I'll then ask you to make and will get you to hold it for a couple of seconds so it can be recorded.

APPENDIX 4 - CLUSTER ANALYSIS PLOTS FOR MILK

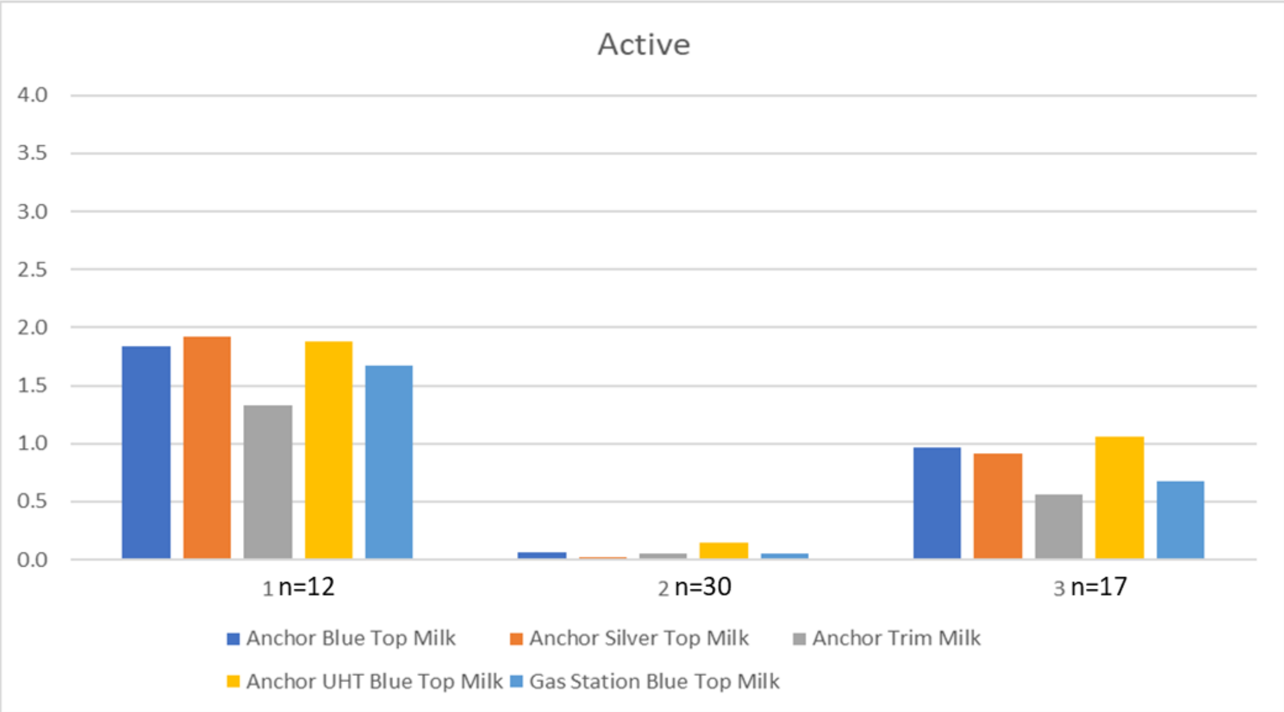

Figure A2: Mean rating of Active (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Active' for milk.

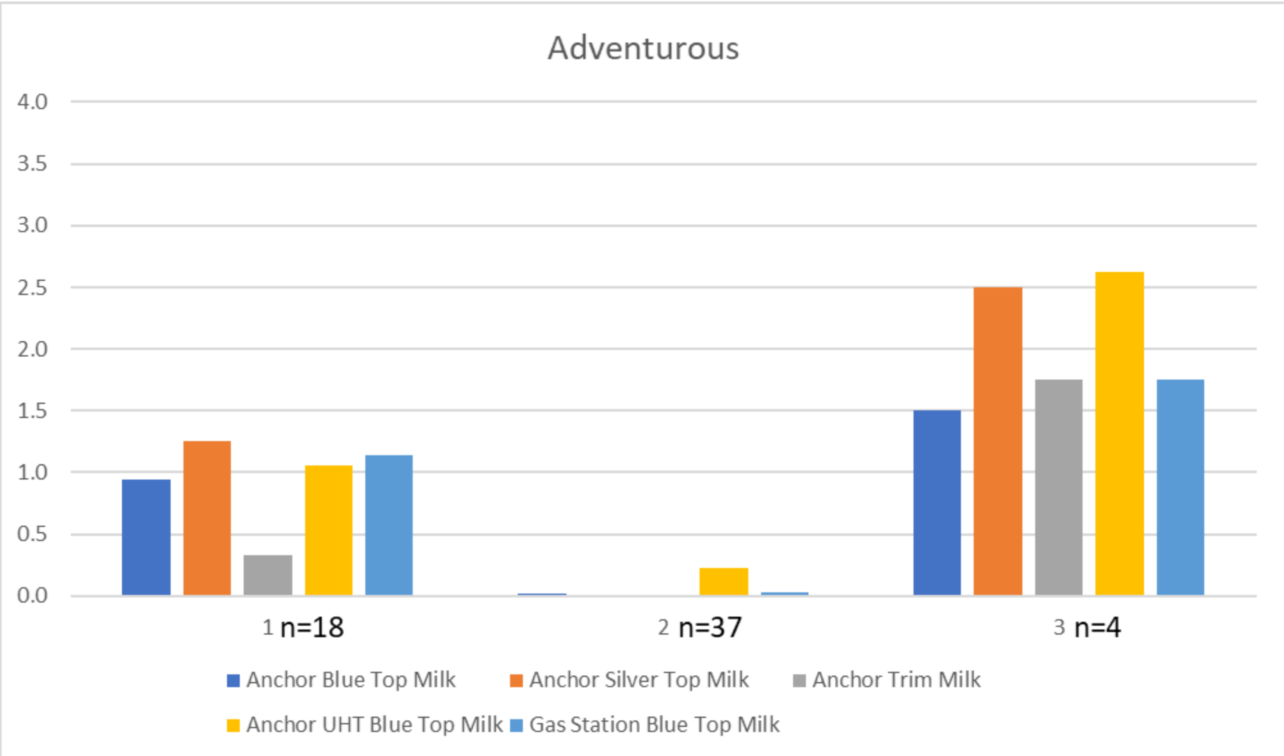

Figure A3: Mean rating of Adventurous (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Adventurous' for milk.

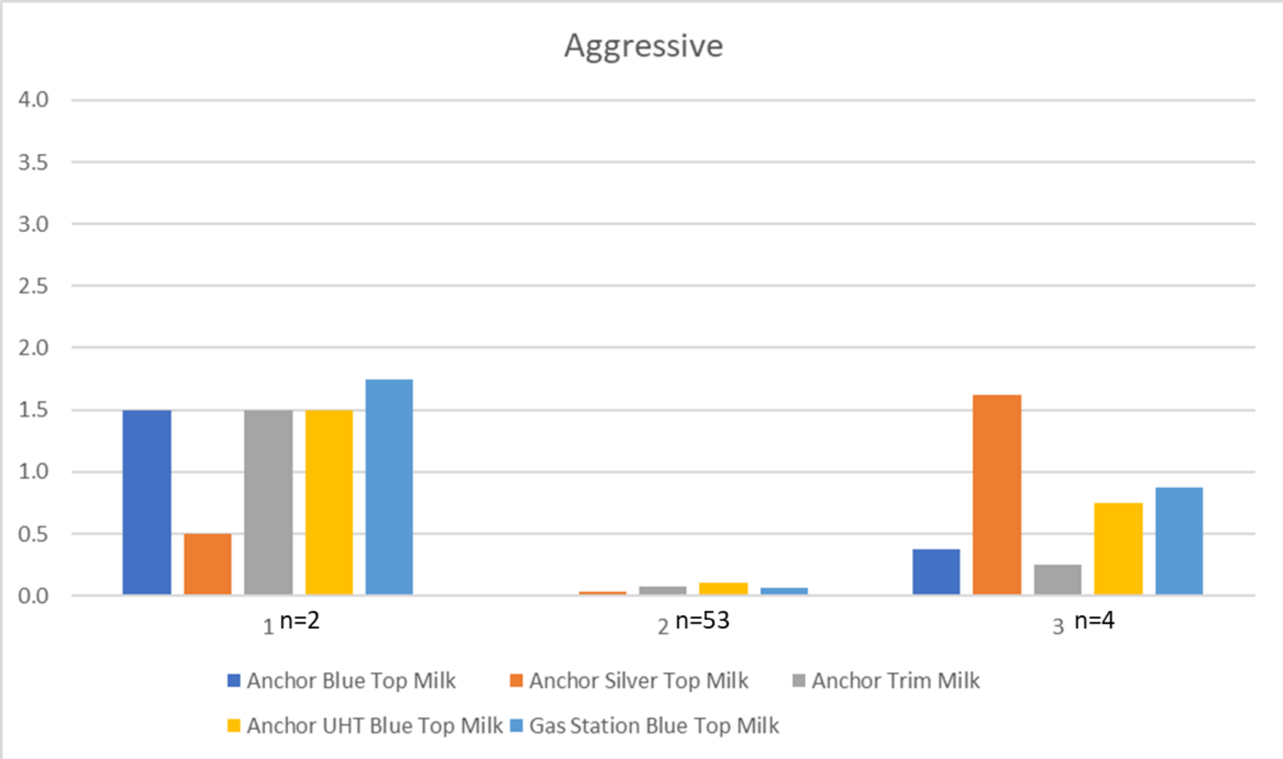

Figure A4: Mean rating of Aggressive (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Aggressive' for milk.

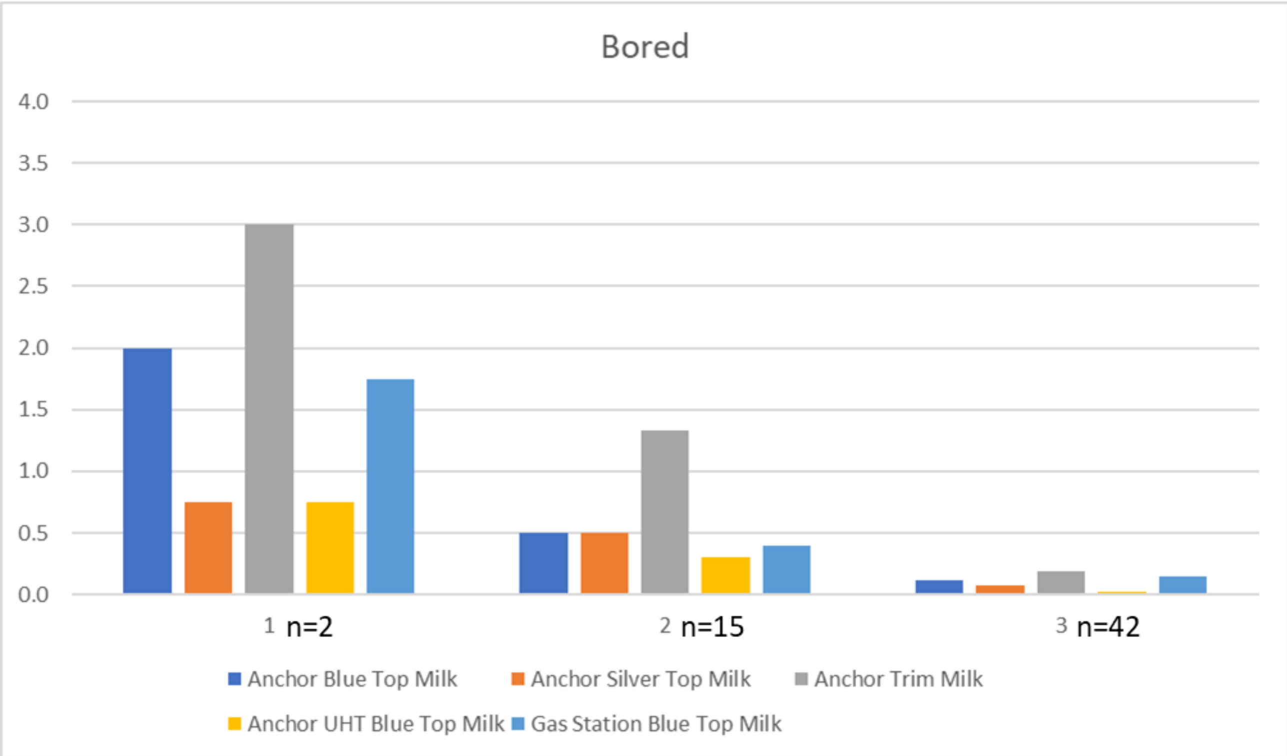

Figure A5: Mean rating of Bored (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Bored' for milk.

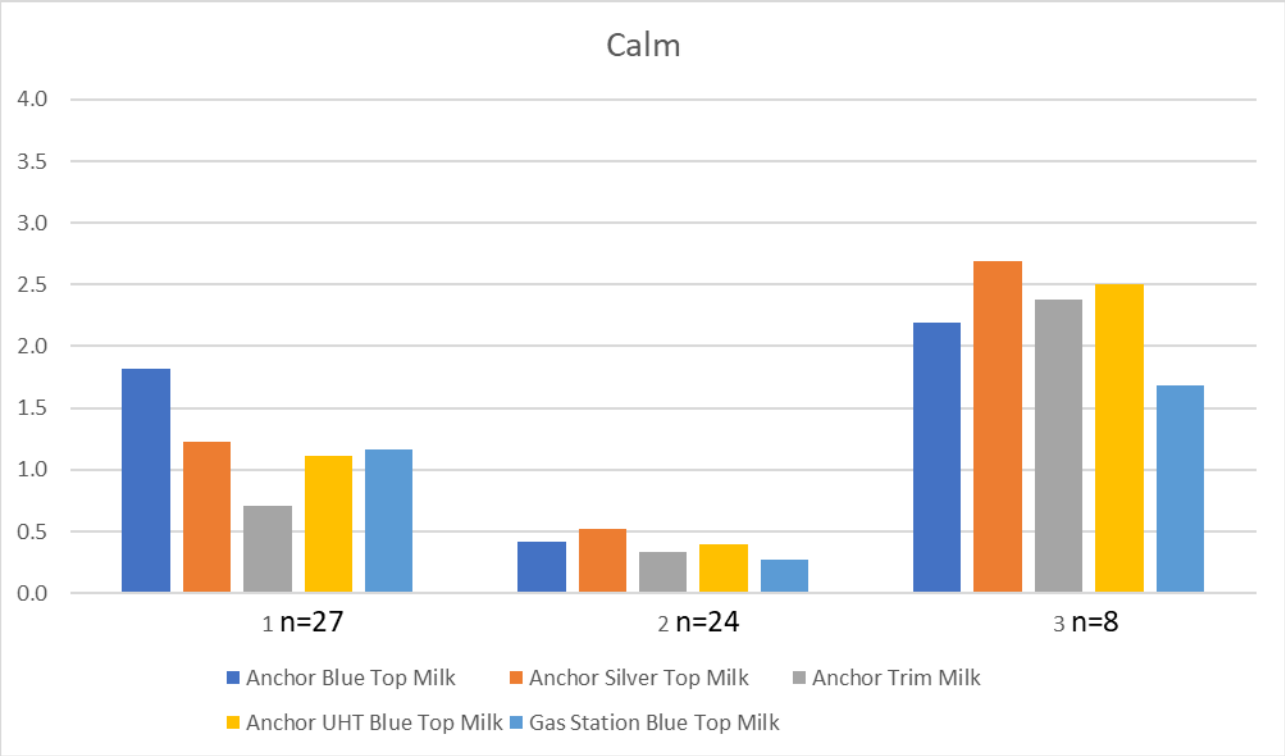

Figure A6: Mean rating of Calm (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Calm' for milk.

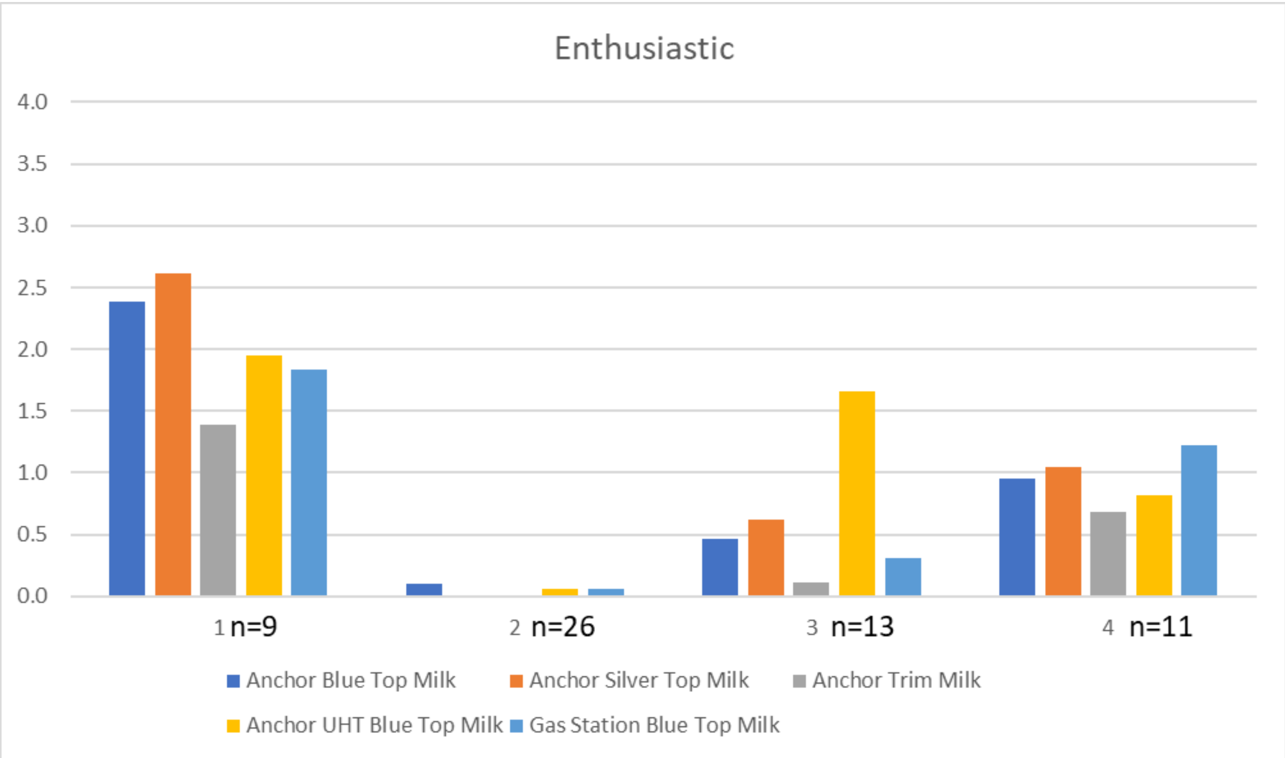

Figure A7: Mean rating of Enthusiastic (on a 0-4 scale) for each milk product for the four clusters of participants grouped by their ratings of 'Enthusiastic' for milk.

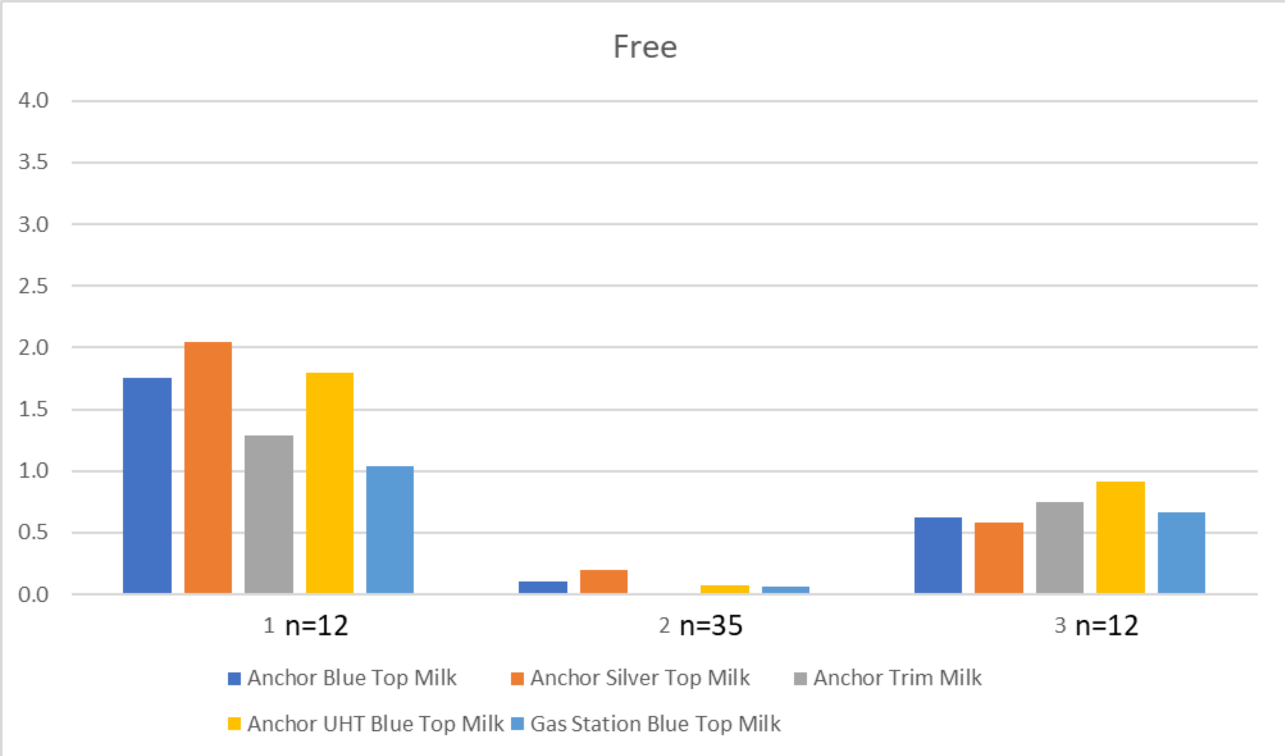

Figure A8: Mean rating of Free (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Free' for milk.

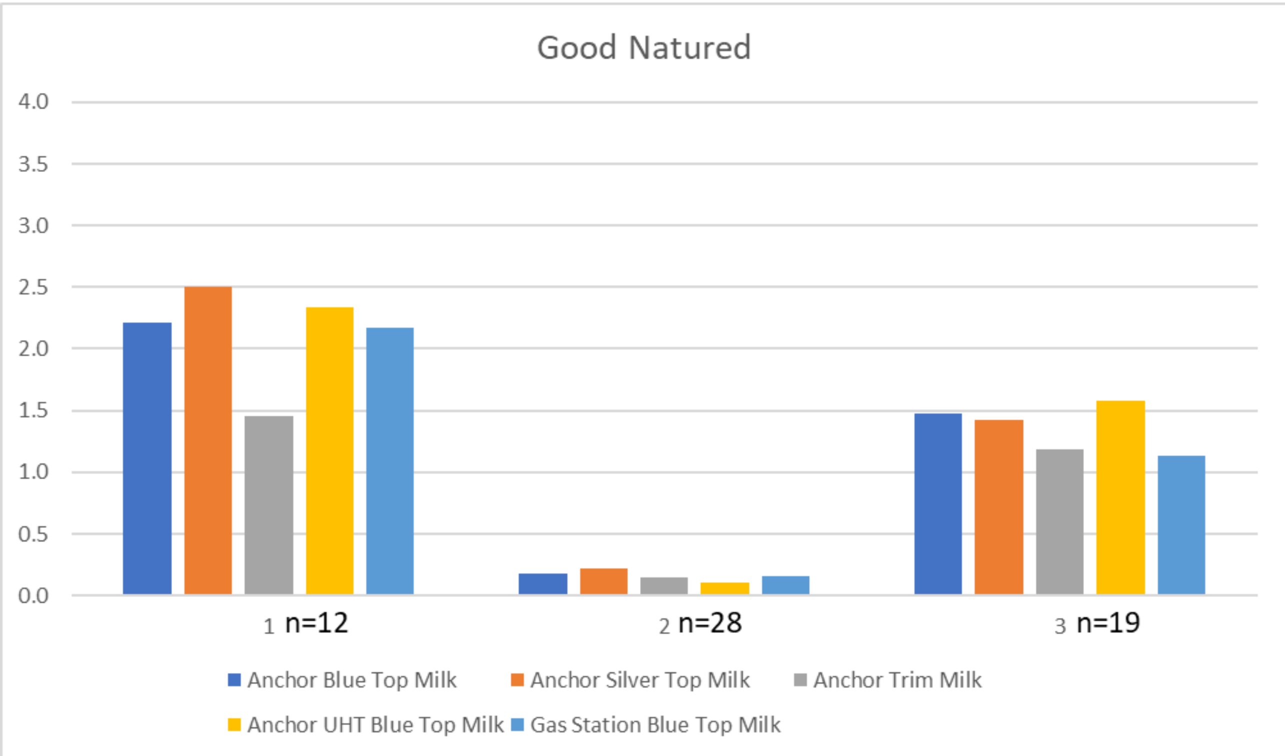

Figure A9: Mean rating of Good natured (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Good natured' for milk.

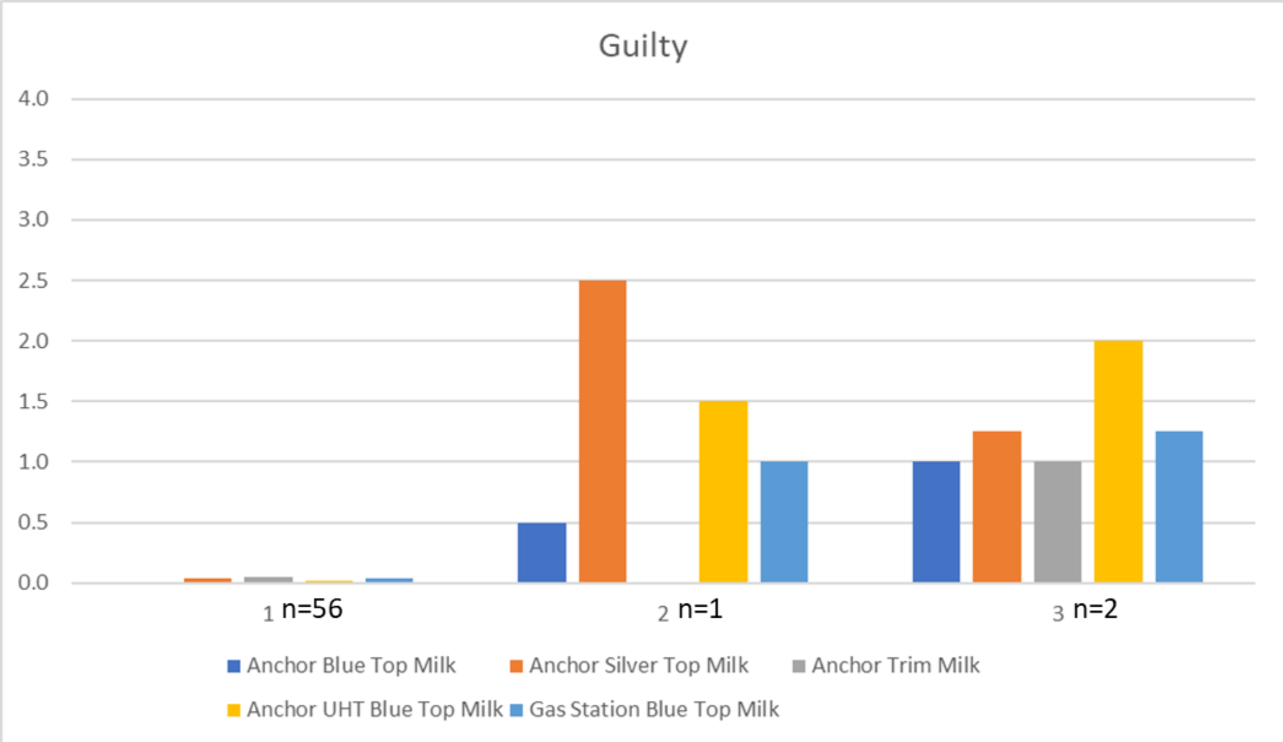

Figure A10: Mean rating of Guilty (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Guilty' for milk.

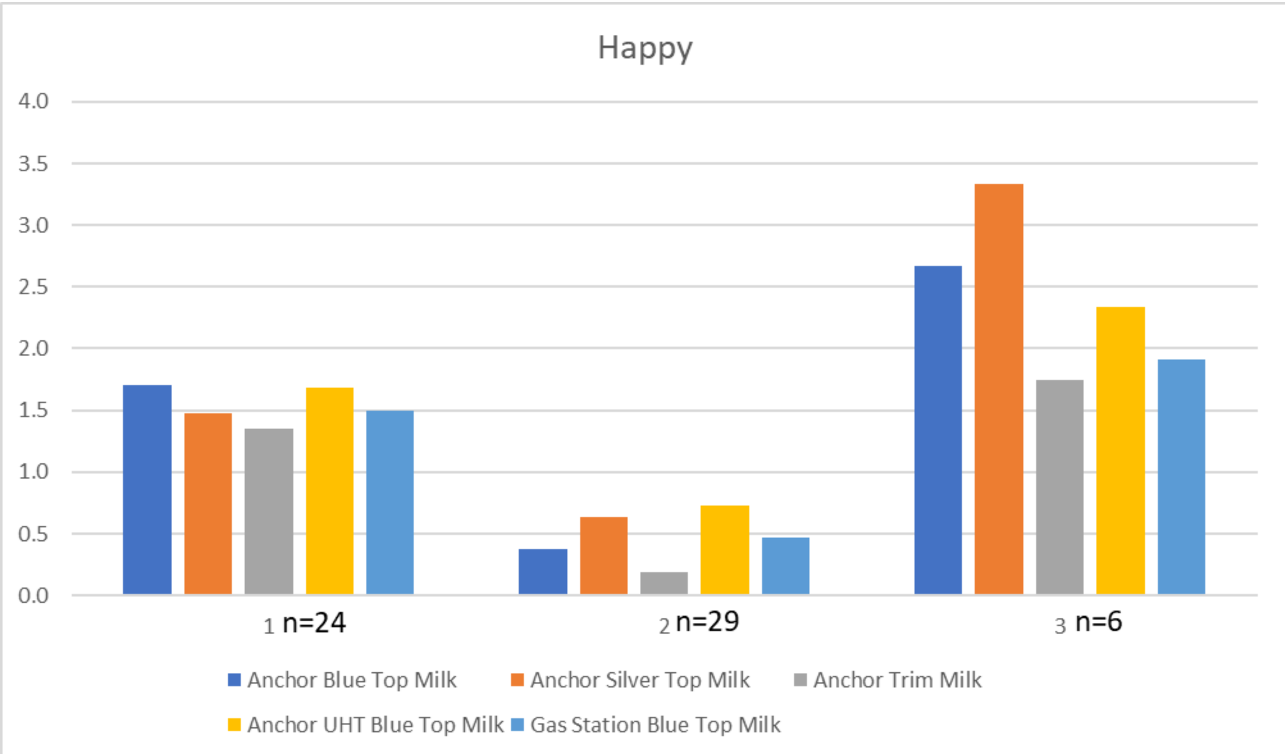

Figure A11: Mean rating of Happy (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Happy' for milk.

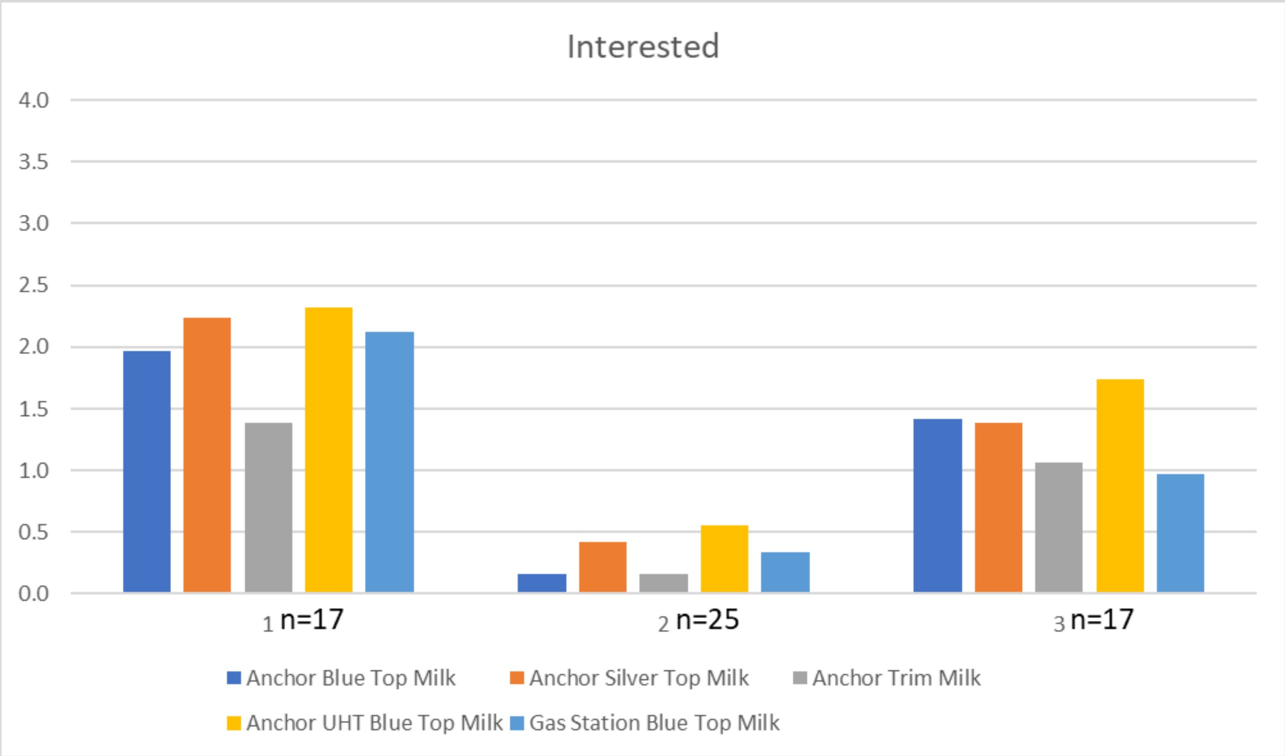

Figure A12: Mean rating of Interested (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Interested' for milk.

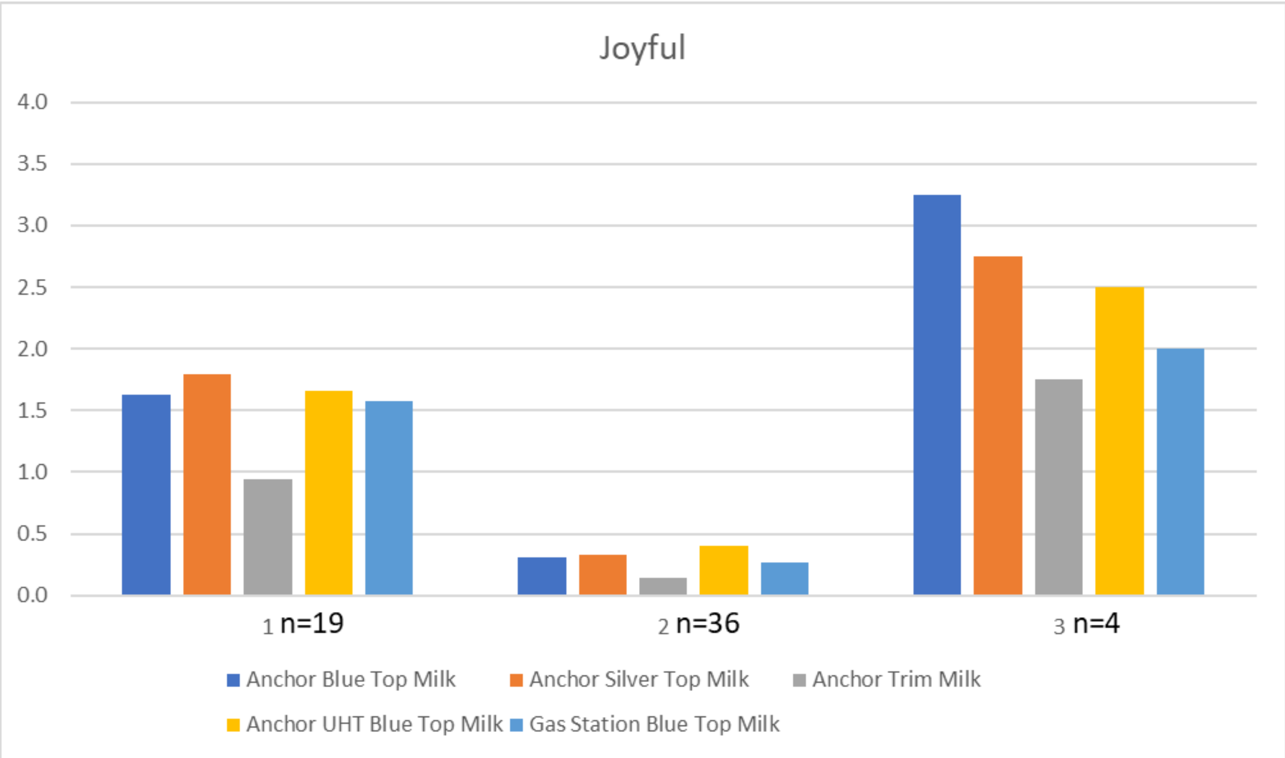

Figure A13: Mean rating of Joyful (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Joyful' for milk.

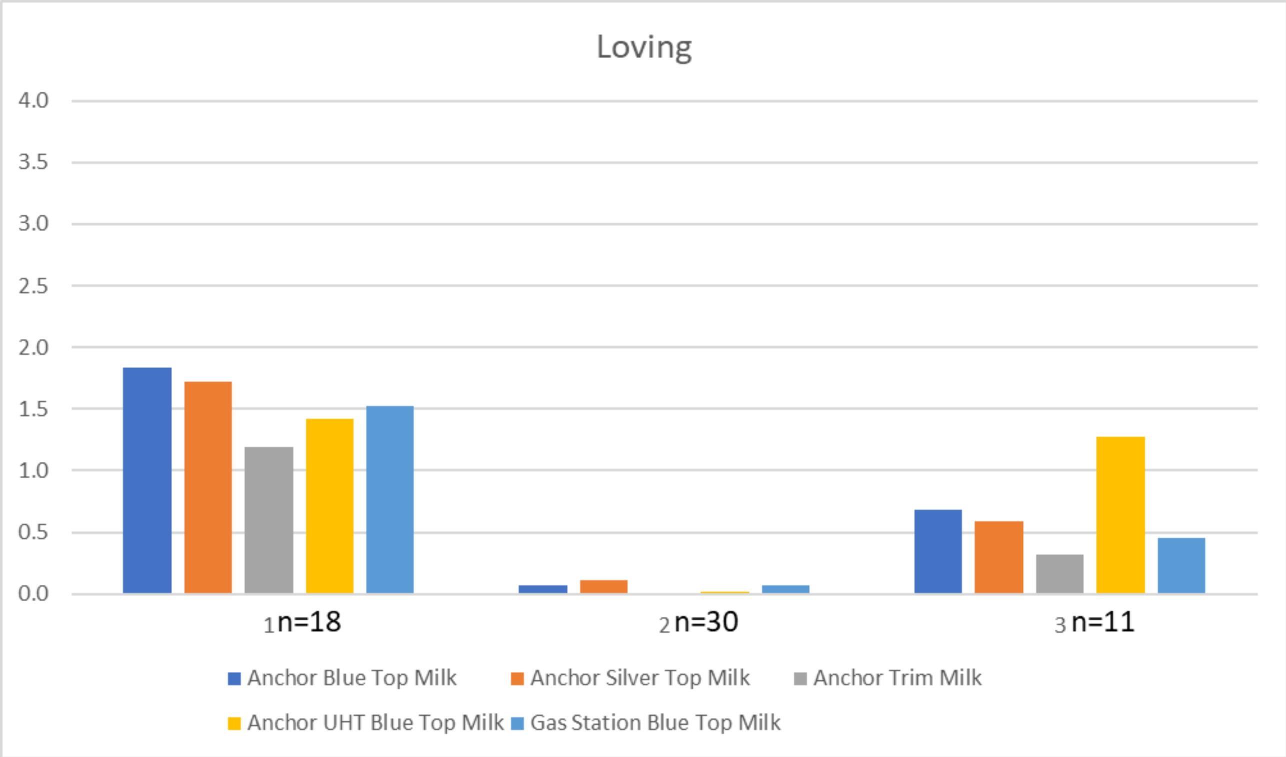

Figure A14: Mean rating of Loving (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Loving' for milk.

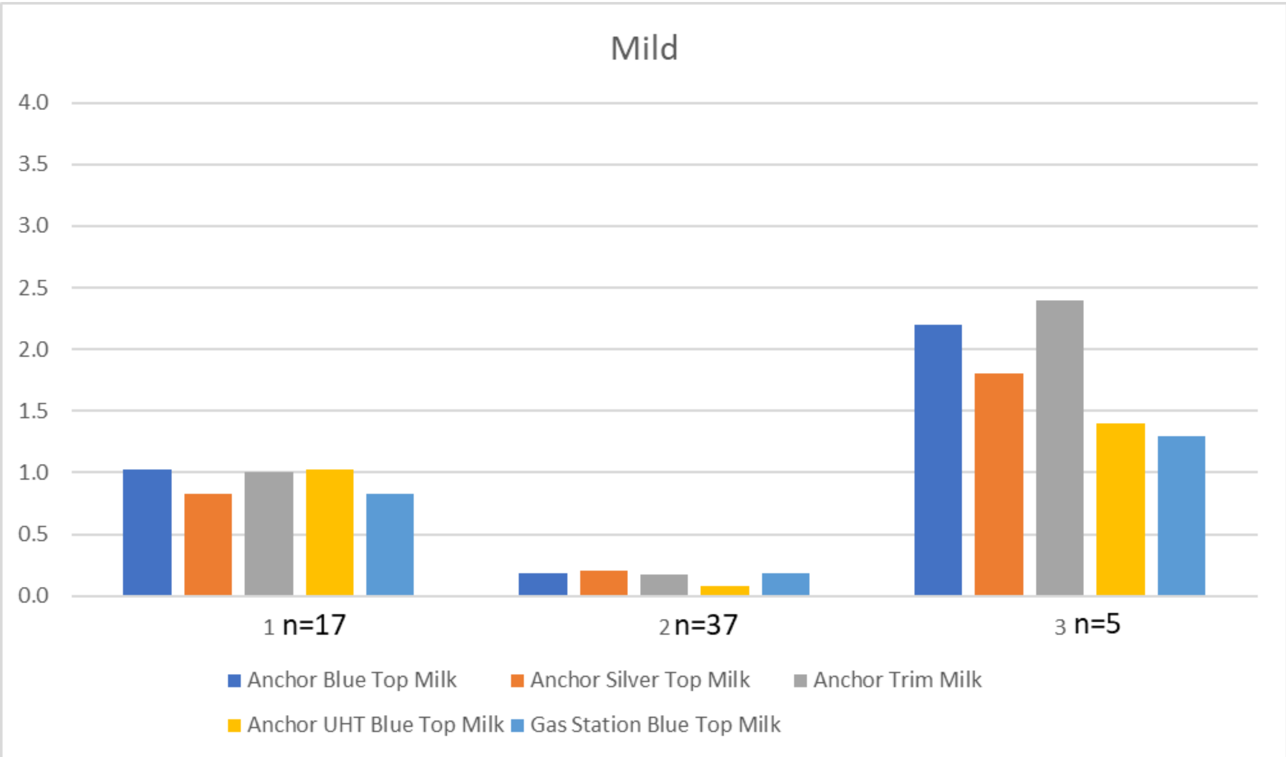

Figure A15: Mean rating of Mild (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Mild' for milk.

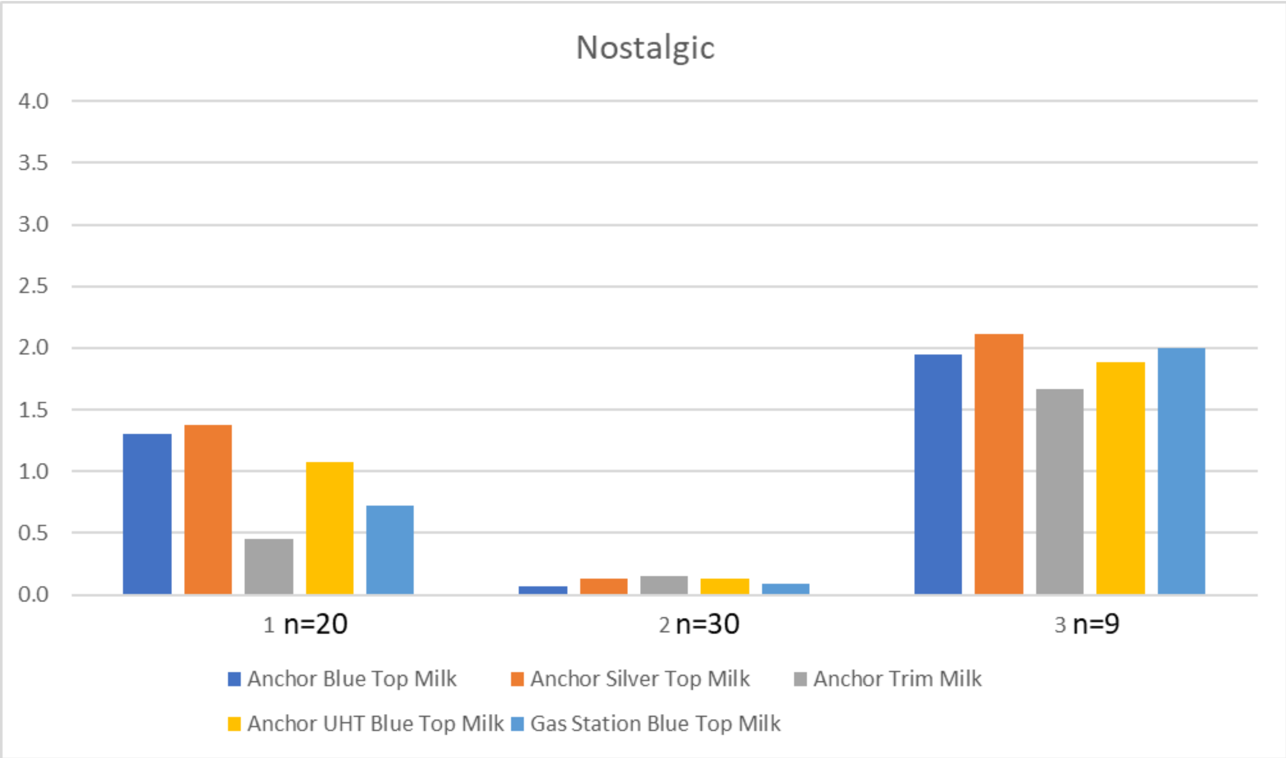

Figure A16: Mean rating of Nostalgic (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Nostalgic' for milk.

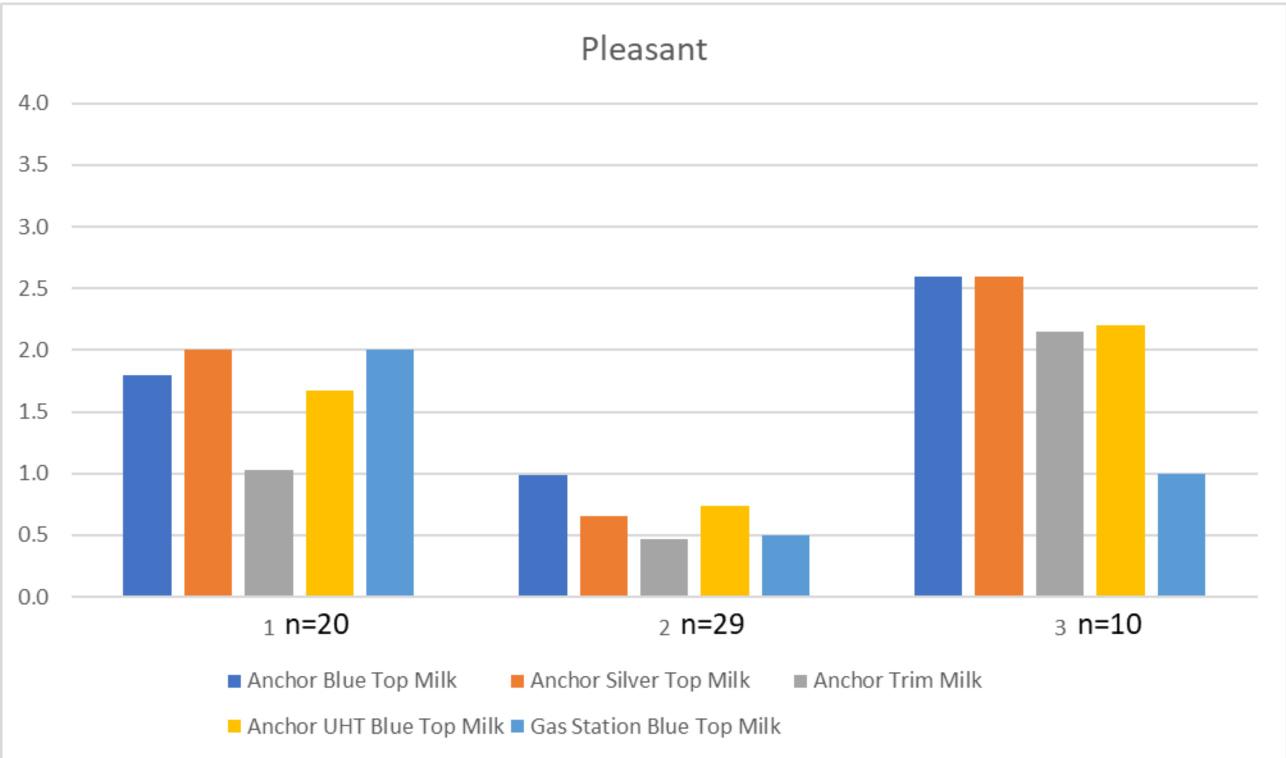

Figure A17: Mean rating of Pleasant (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Pleasant' for milk.

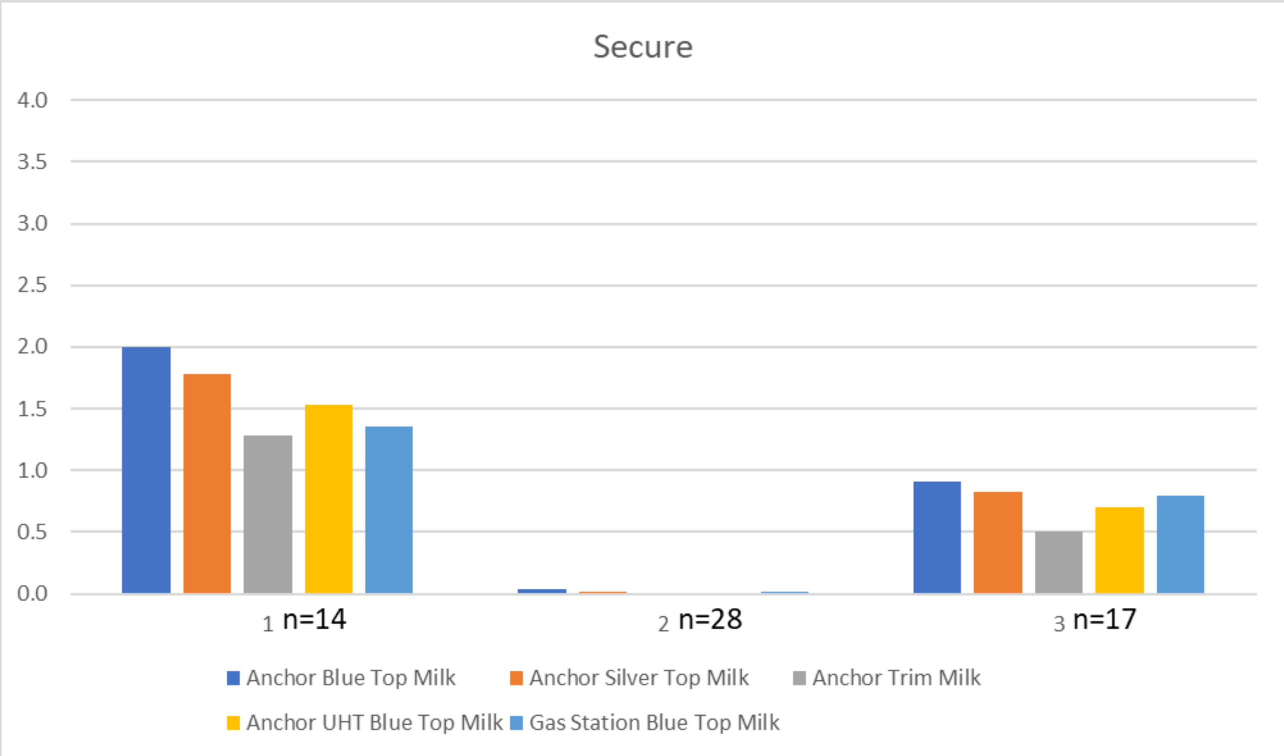

Figure A18: Mean rating of Secure (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Secure' for milk.

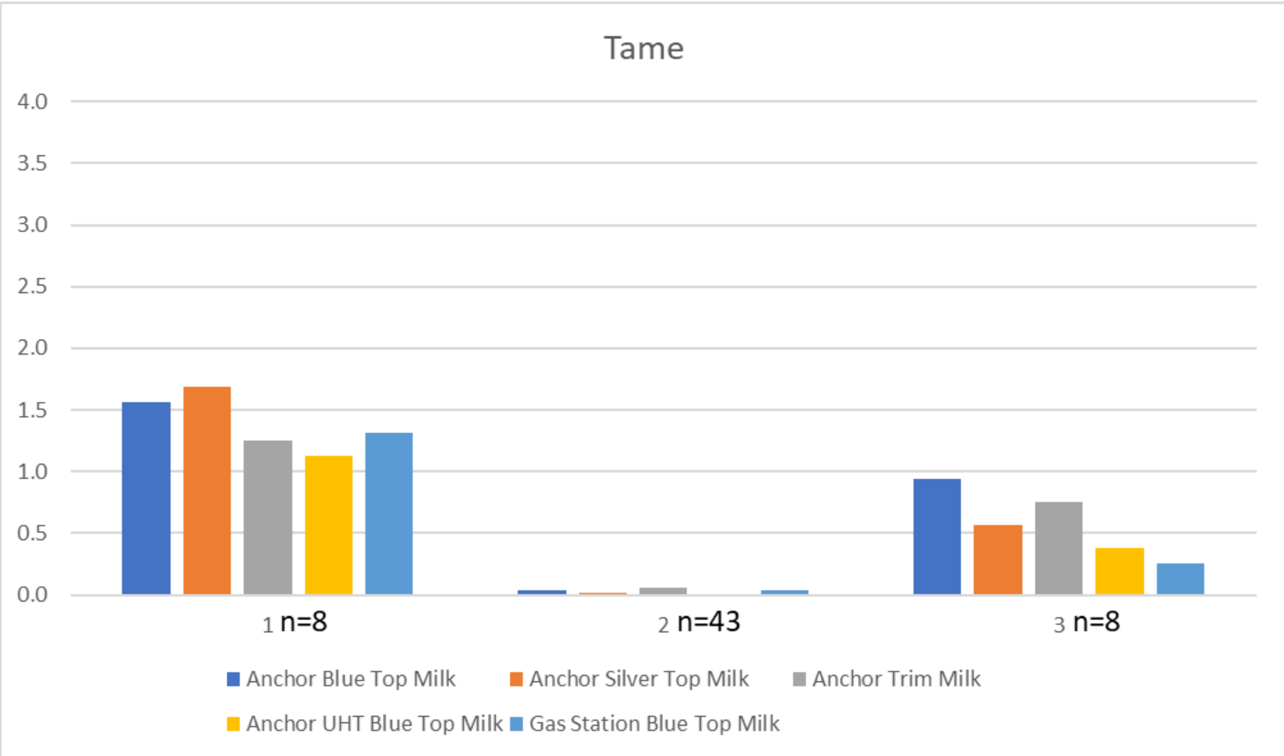

Figure A19: Mean rating of Tame (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Tame' for milk.

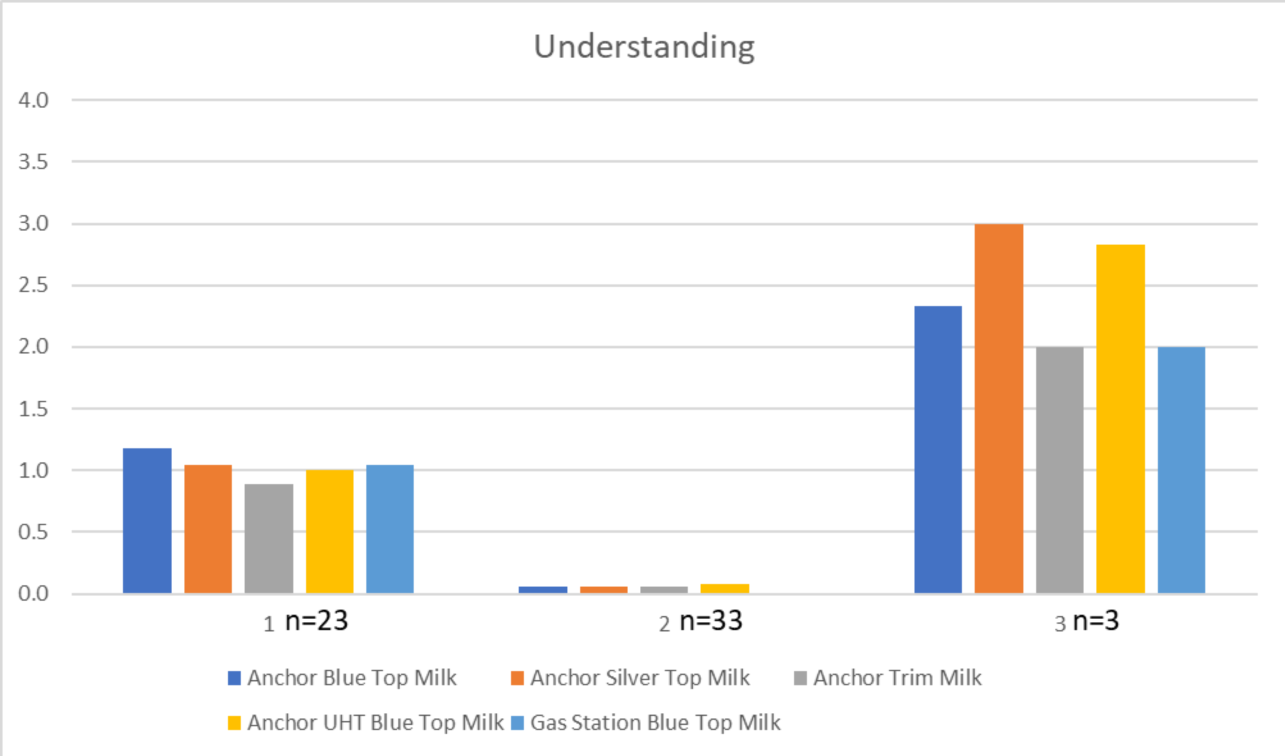

Figure A20: Mean rating of Understanding (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Understanding' for milk.

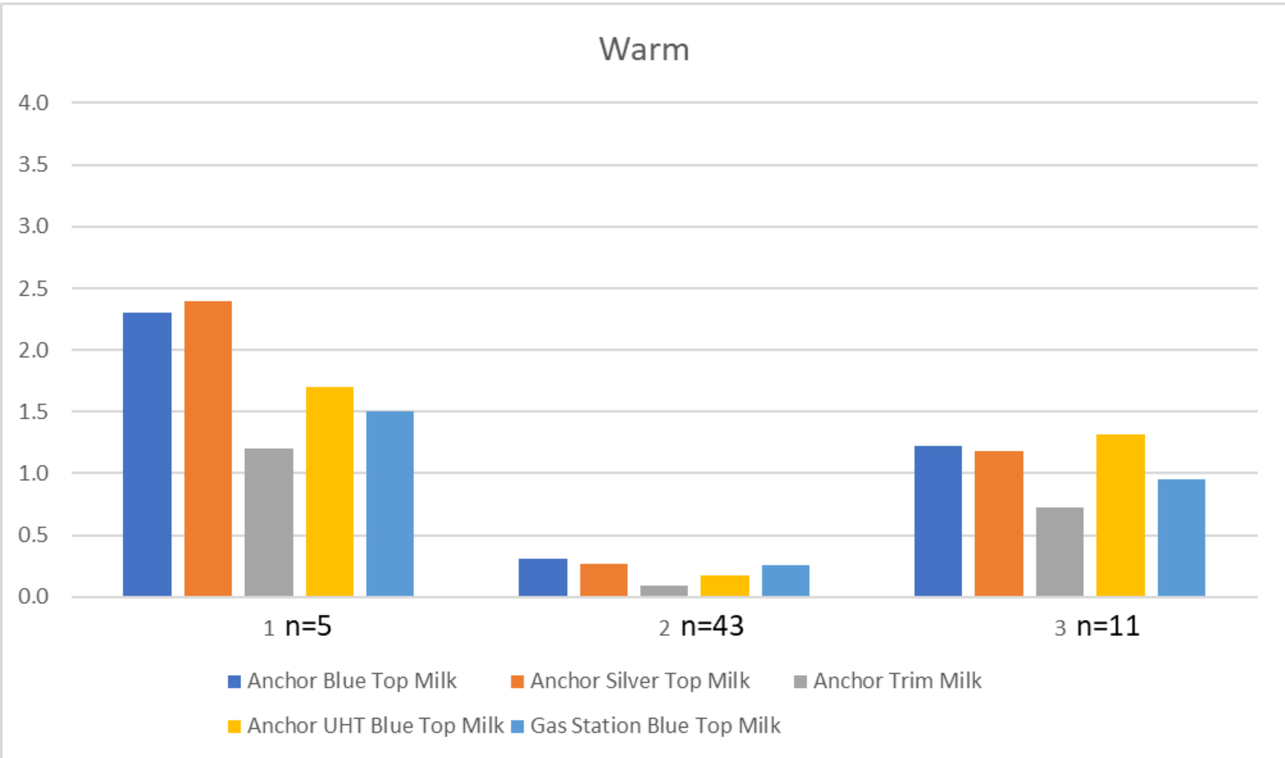

Figure A21: Mean rating of Warm (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Warm' for milk.

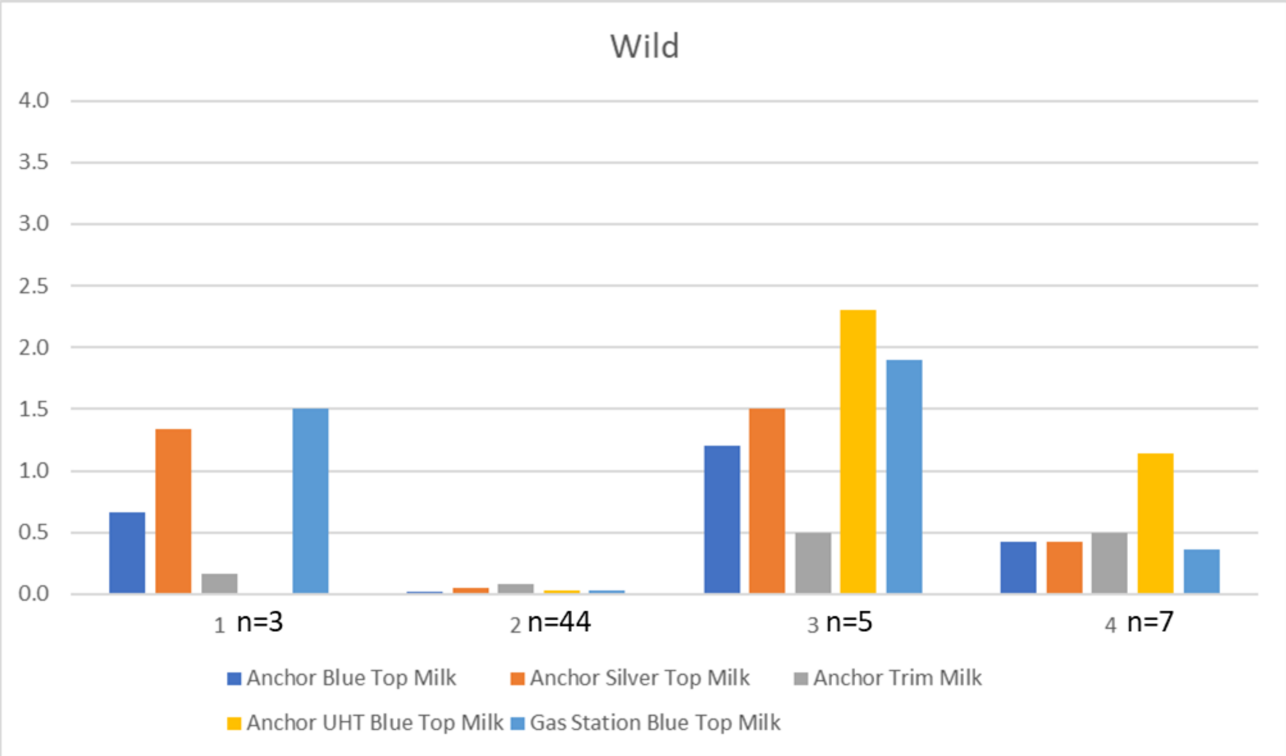

Figure A22: Mean rating of Wild (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Wild' for milk.

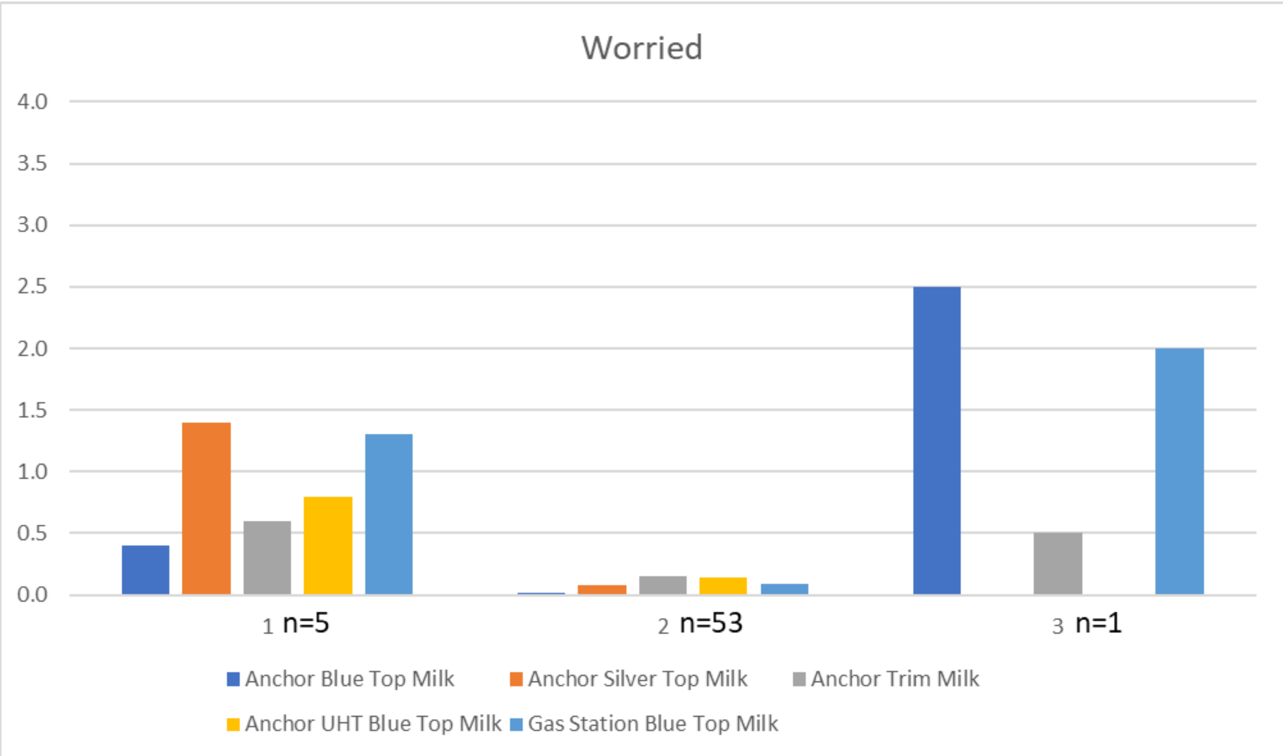

Figure A23: Mean rating of Worried (on a 0-4 scale) for each milk product for the three clusters of participants grouped by their ratings of 'Worried' for milk.

APPENDIX 5 - CLUSTER ANALYSIS PLOTS FOR YOGHURT

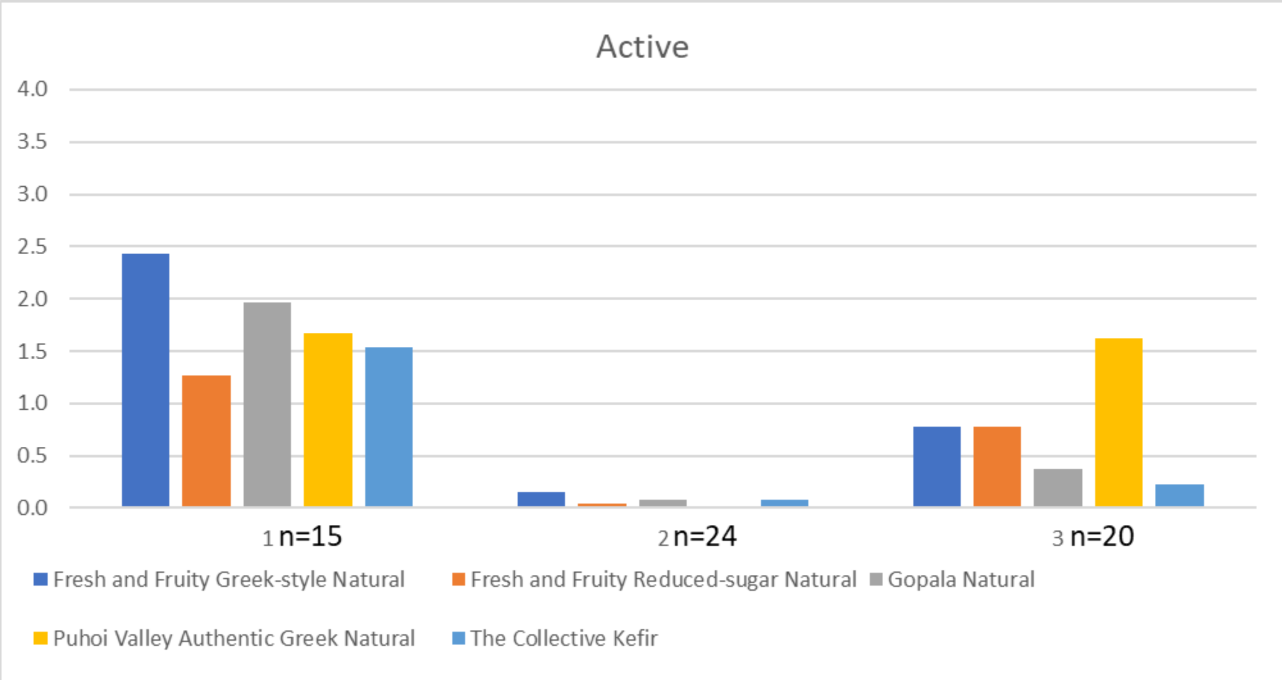

Figure A24: Mean rating of Active (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Active' for yoghurt.

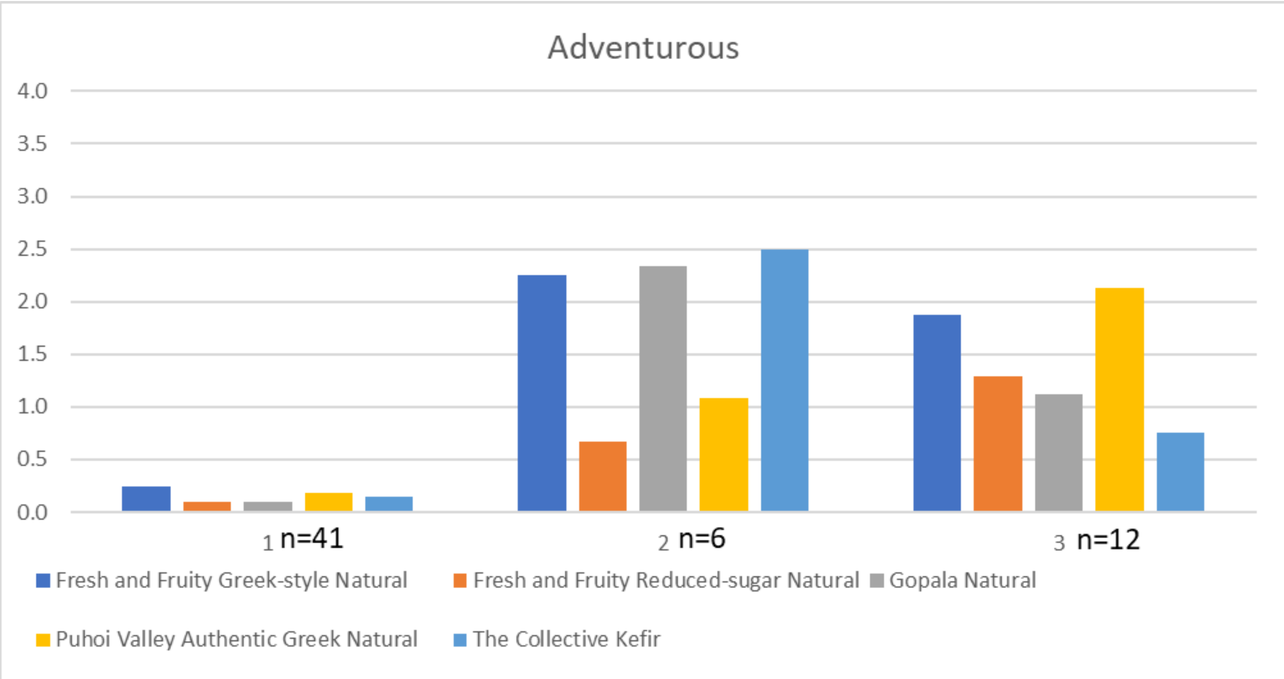

Figure A25: Mean rating of Adventurous (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Adventurous' for yoghurt.

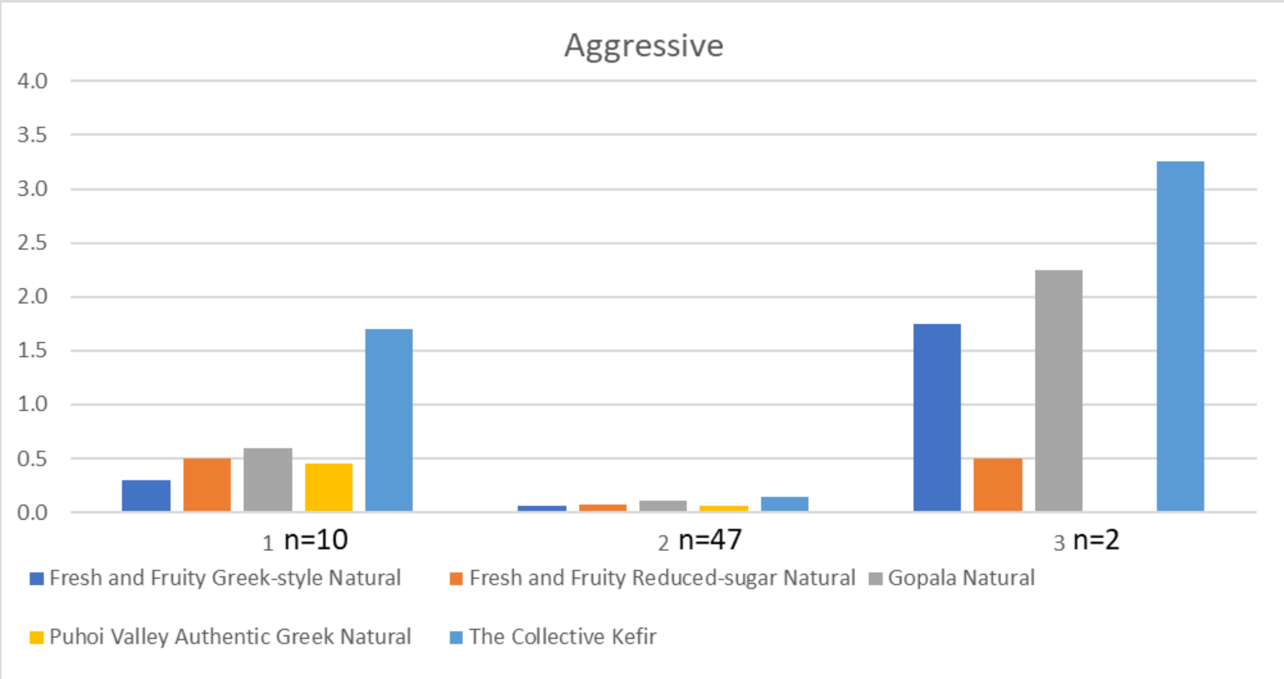

Figure A26: Mean rating of Aggressive (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Aggressive' for yoghurt.

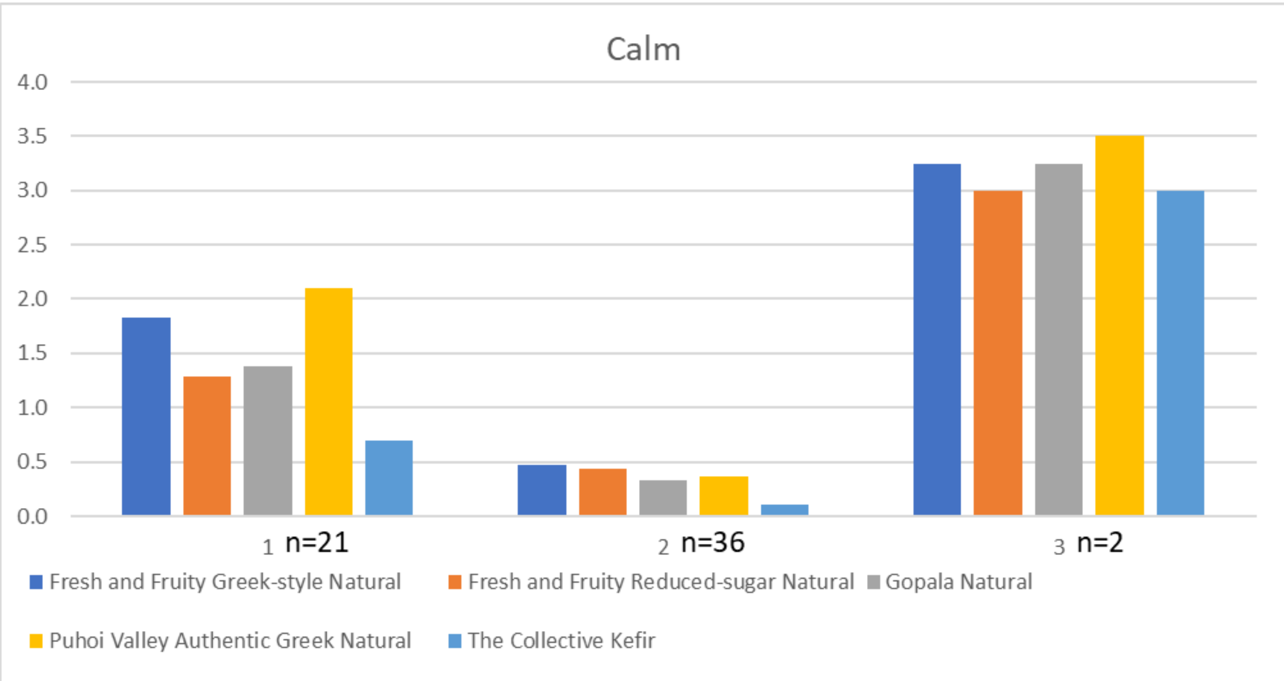

Figure A27: Mean rating of Calm (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Calm' for yoghurt.

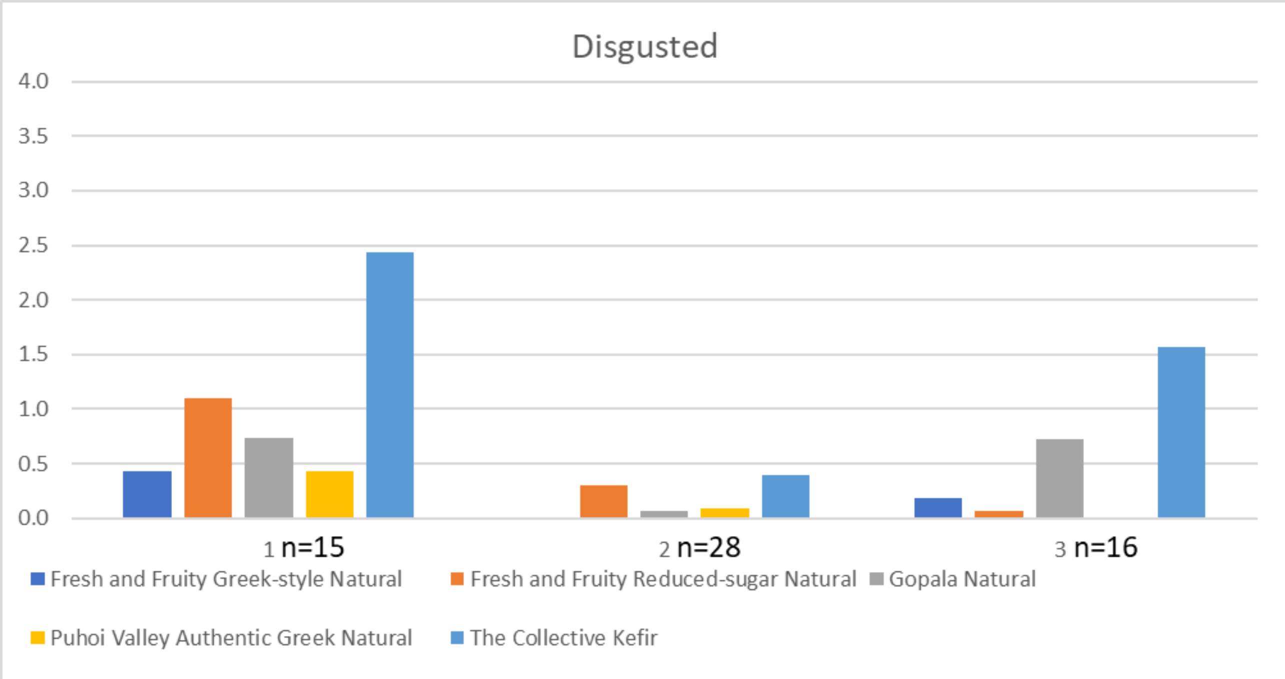

Figure A28: Mean rating of Disgusted (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Disgusted' for yoghurt.

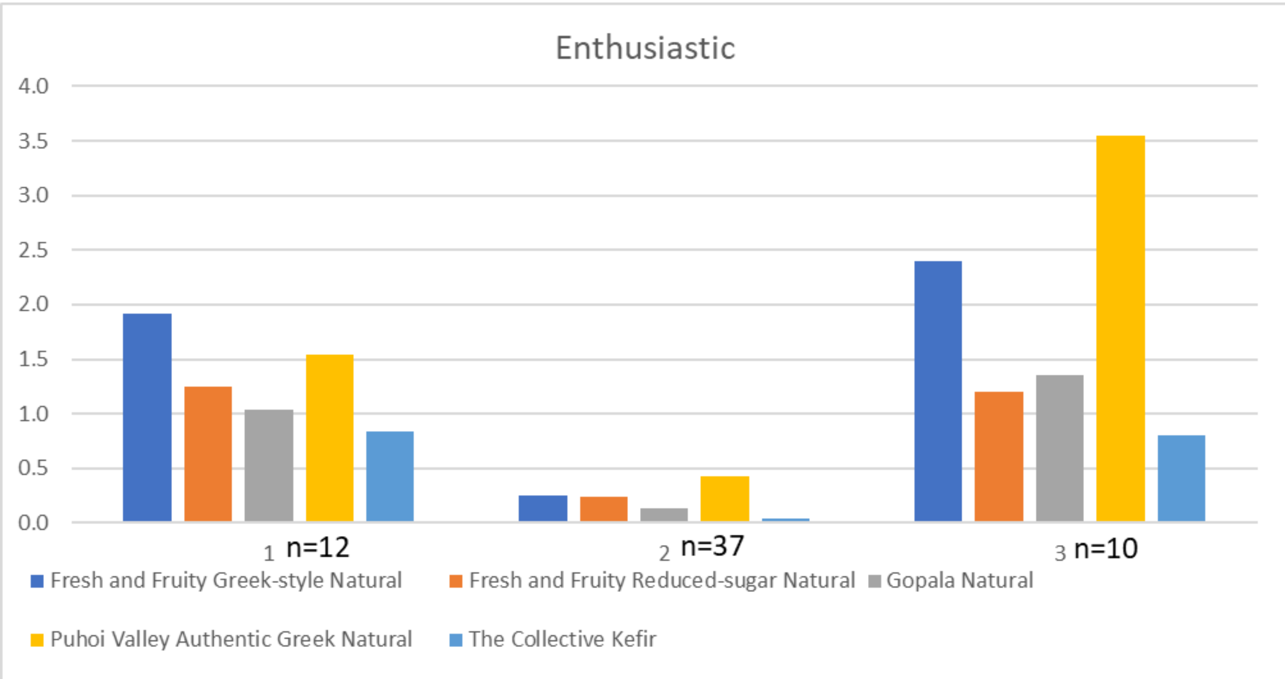

Figure A29: Mean rating of Enthusiastic (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Enthusiastic' for yoghurt.

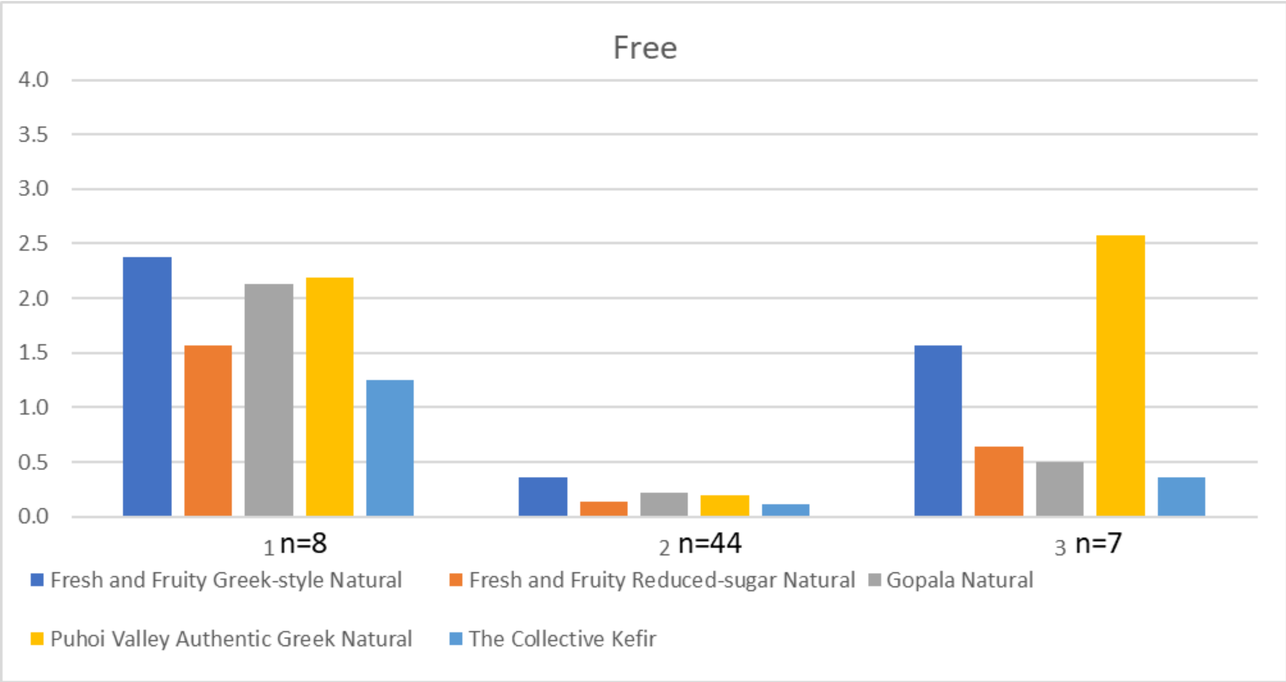

Figure A30: Mean rating of Free (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Free' for yoghurt.

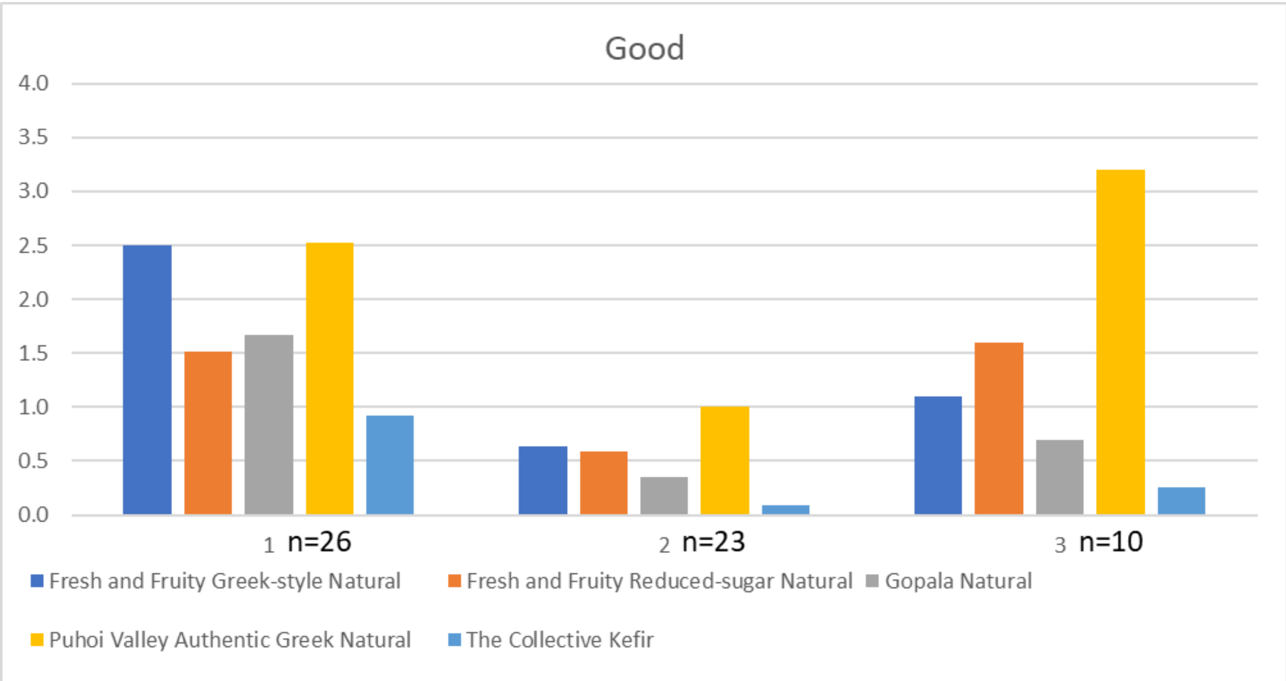

Figure A31: Mean rating of Good (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Good' for yoghurt.

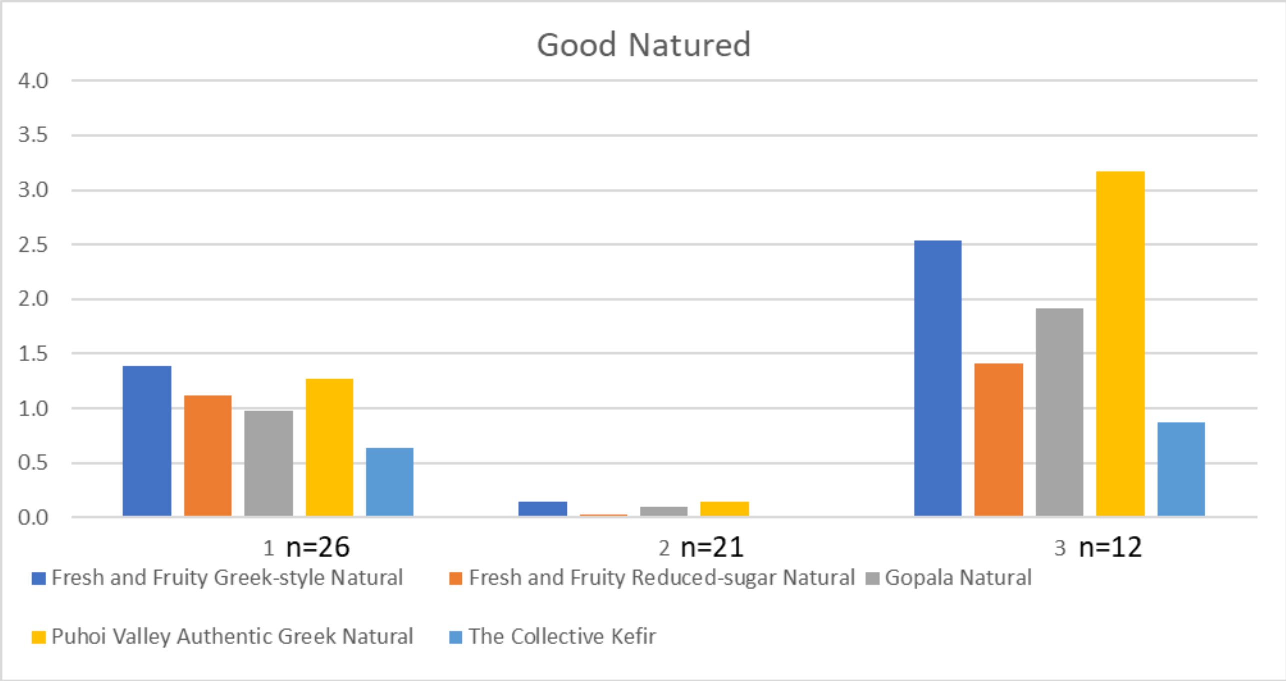

Figure A32: Mean rating of Good natured (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Good natured' for yoghurt.

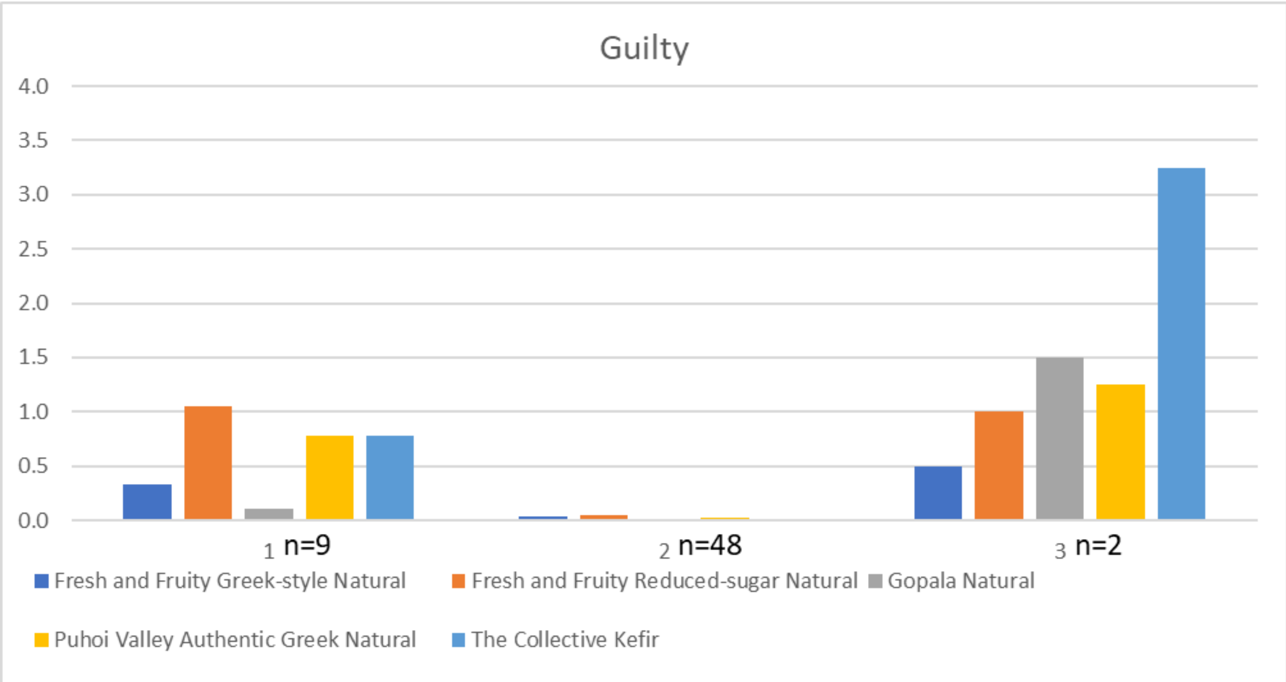

Figure A33: Mean rating of Guilty (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Guilty' for yoghurt.

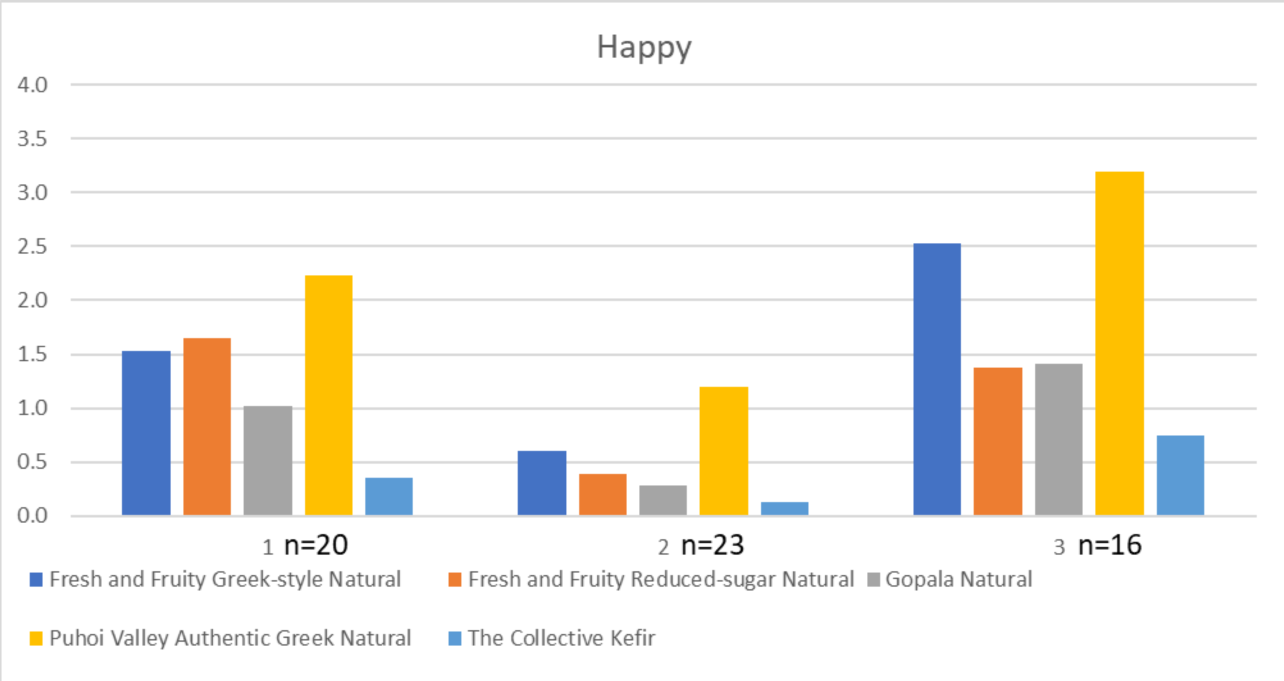

Figure A34: Mean rating of Happy (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Happy' for yoghurt.

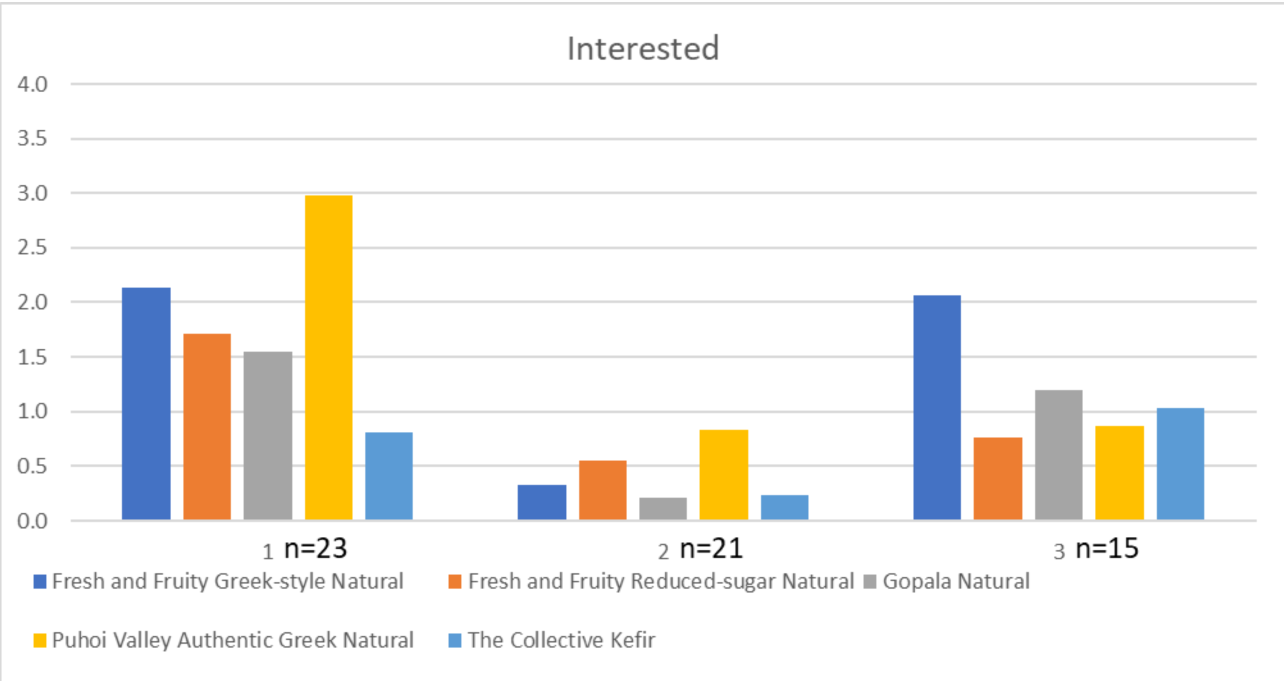

Figure A35: Mean rating of Interested (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Interested' for yoghurt.

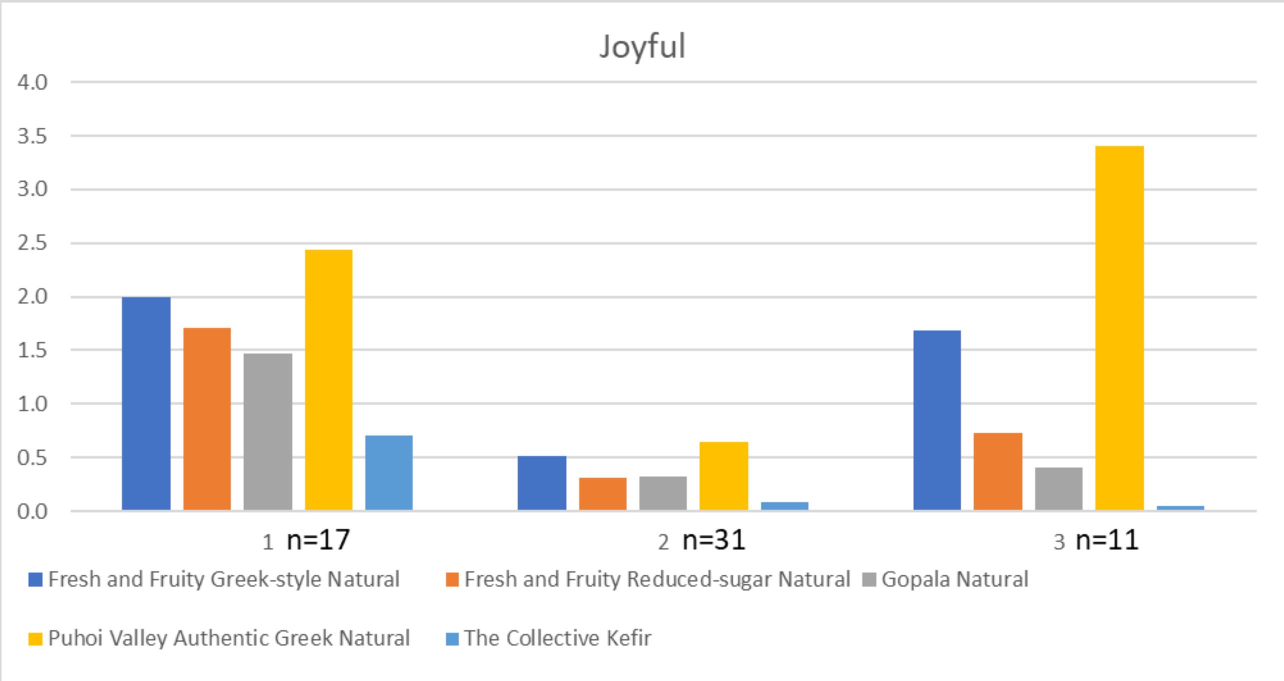

Figure A36: Mean rating of Joyful (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Joyful' for yoghurt.

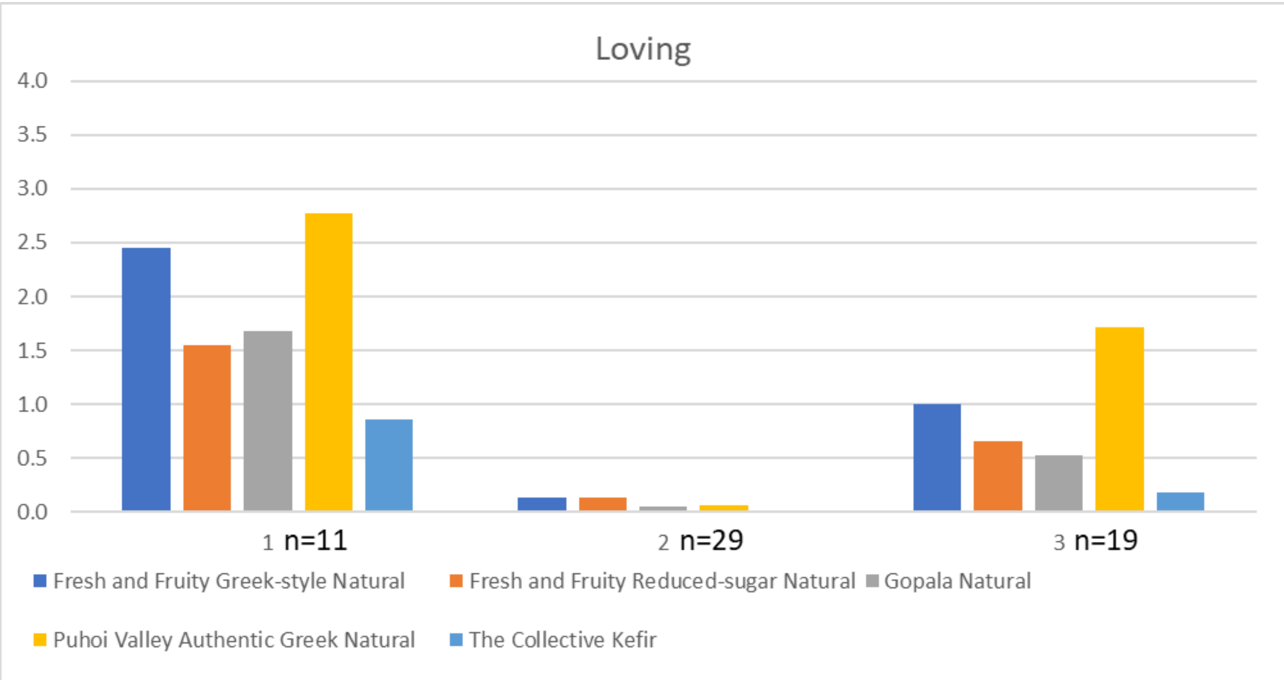

Figure A37: Mean rating of Loving (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Loving' for yoghurt.

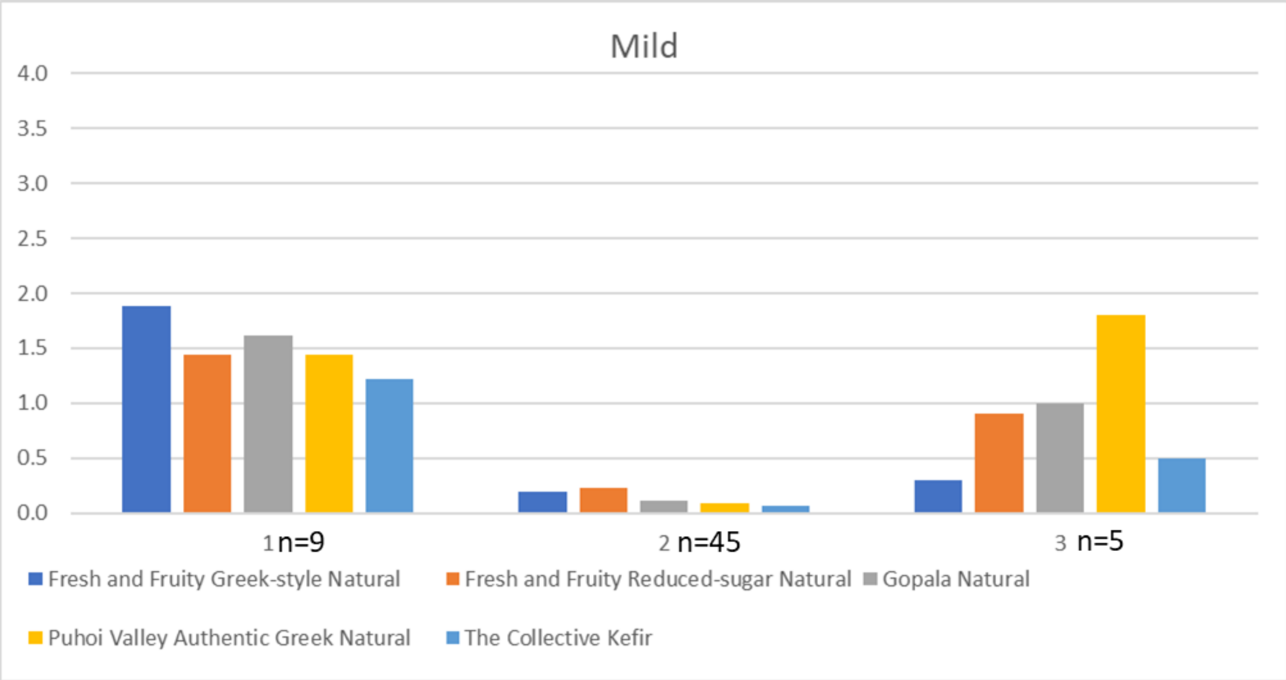

Figure A38: Mean rating of Mild (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Mild' for yoghurt.

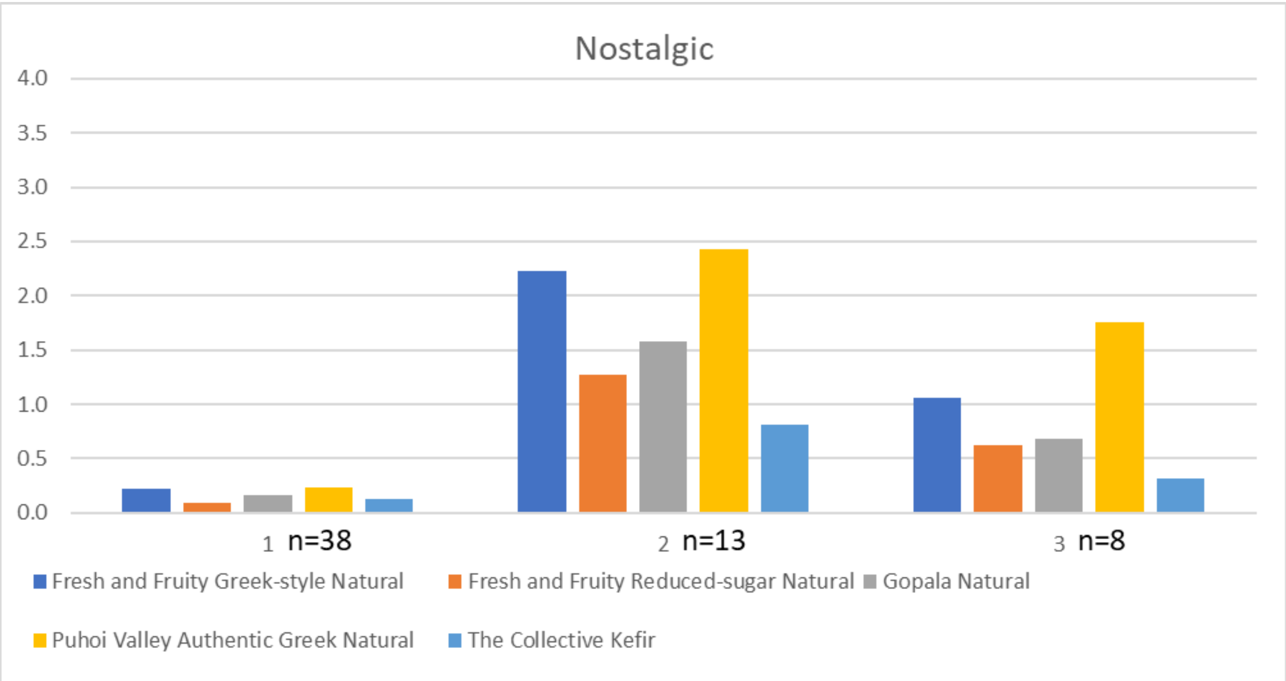

Figure A39: Mean rating of Nostalgic (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Nostalgic' for yoghurt.

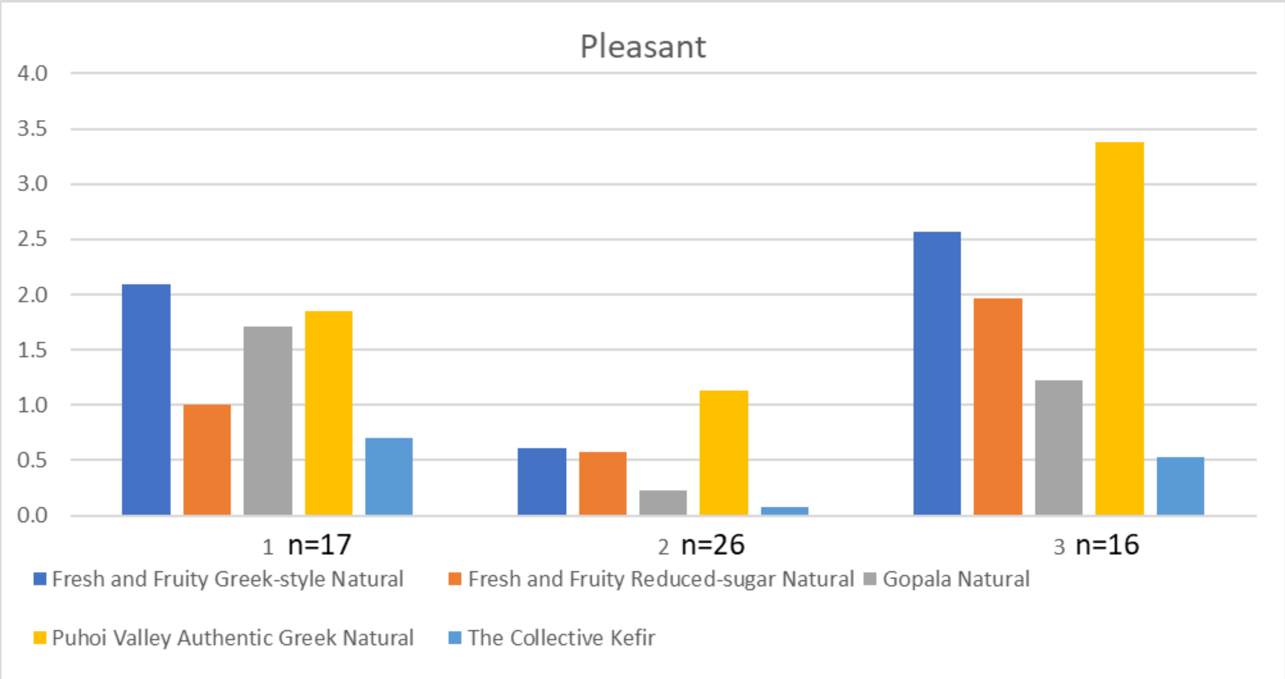

Figure A40: Mean rating of Pleasant (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Pleasant' for yoghurt.

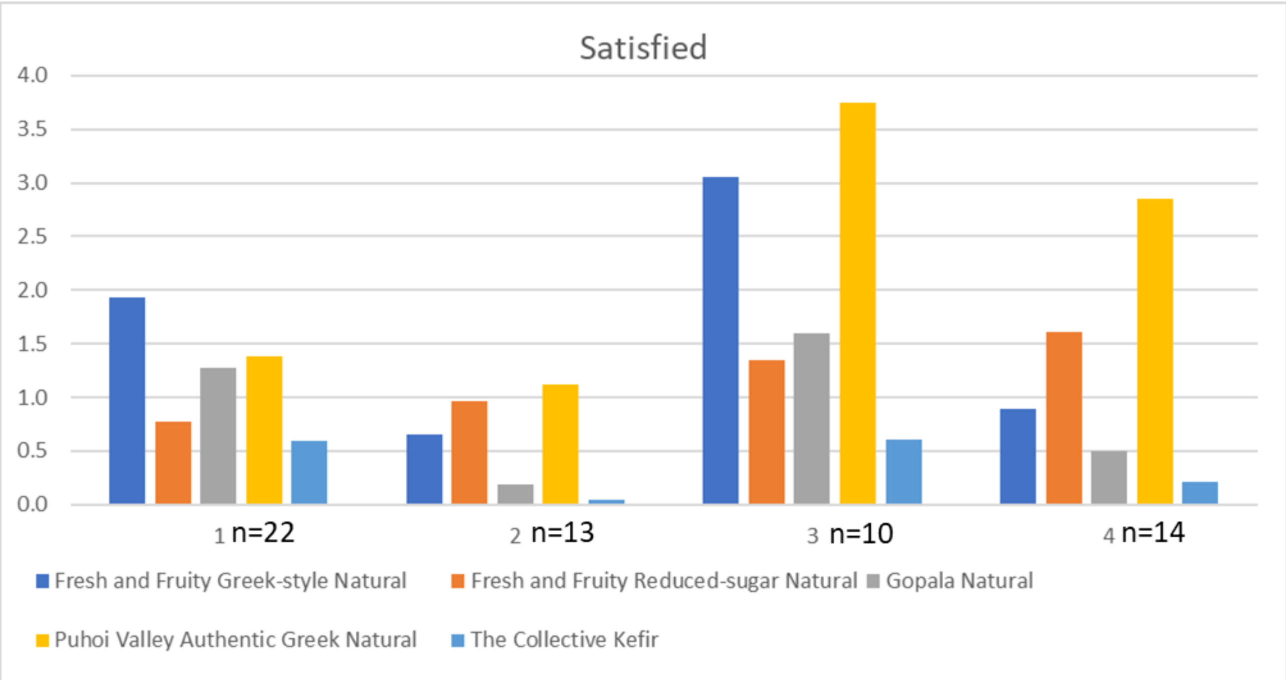

Figure A41: Mean rating of Satisfied (on a 0-4 scale) for each yoghurt product for the four clusters of participants grouped by their ratings of 'Satisfied' for yoghurt.

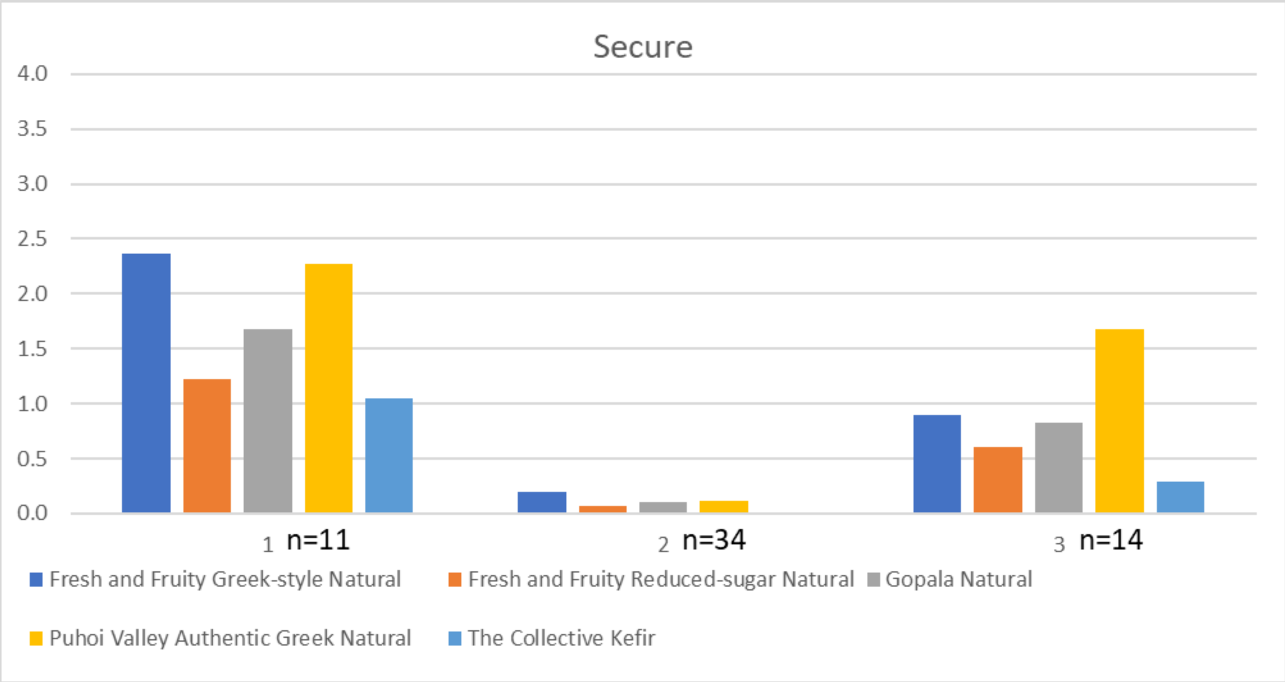

Figure A42: Mean rating of Secure (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Secure' for yoghurt.

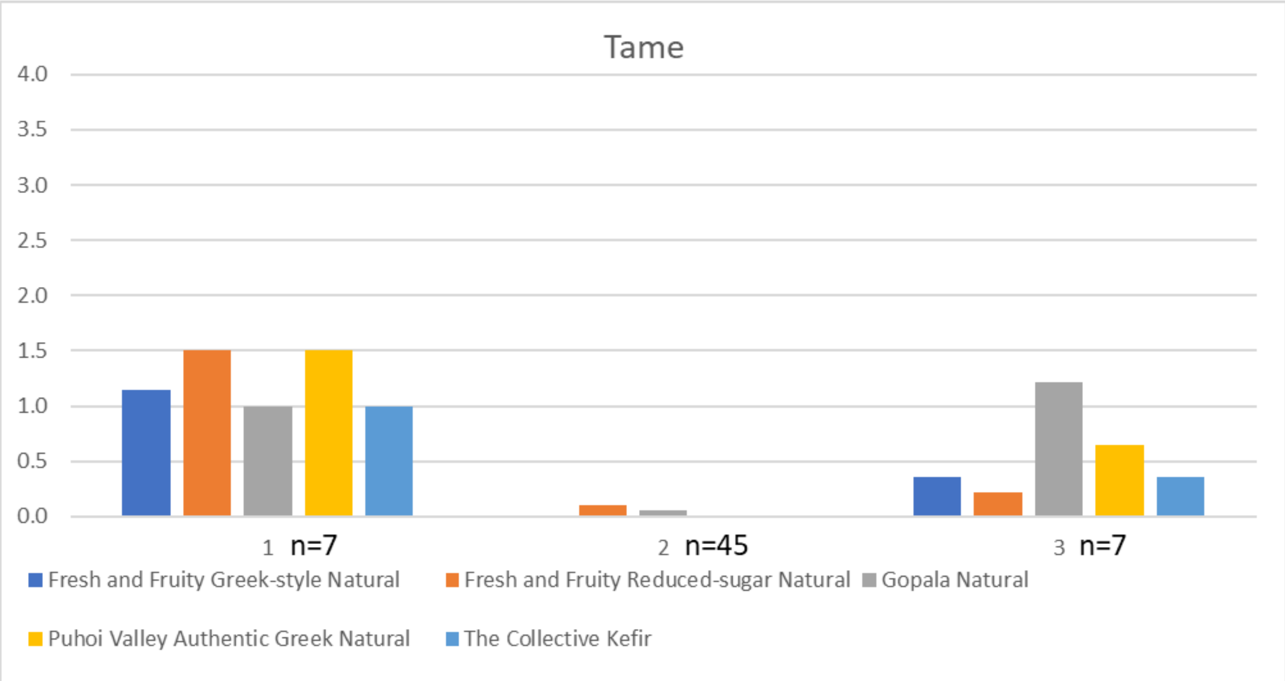

Figure A43: Mean rating of Tame (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Tame' for yoghurt.

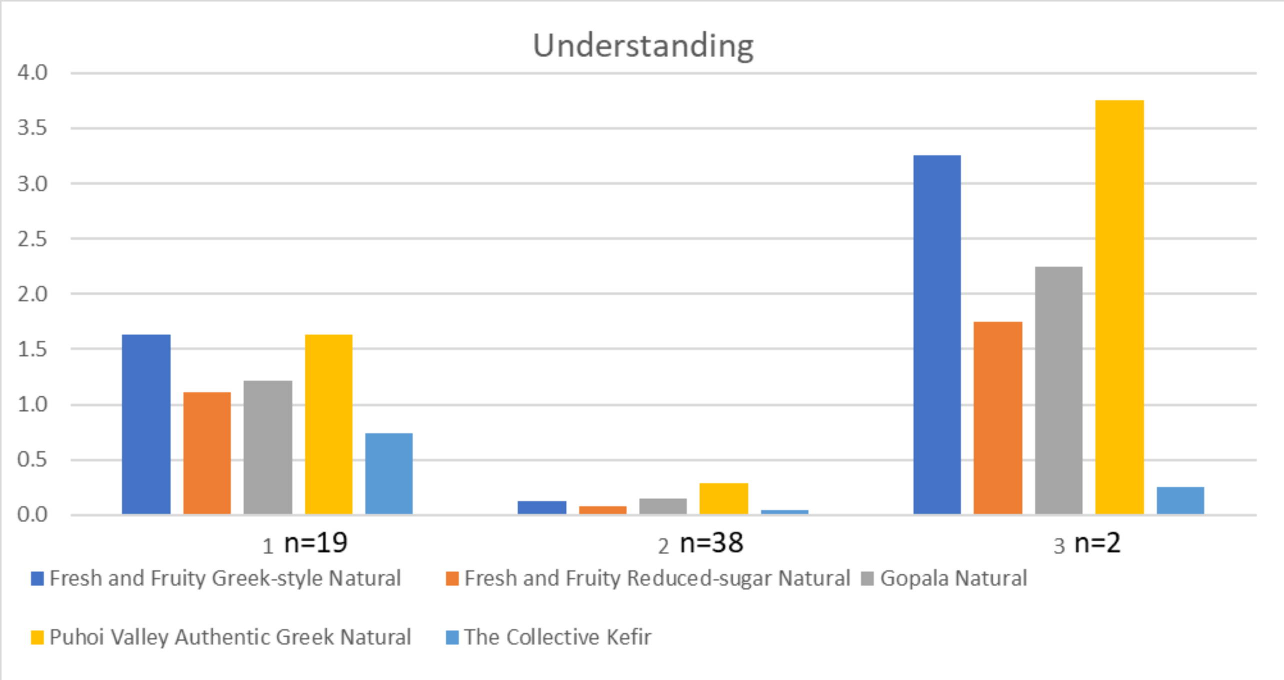

Figure A44: Mean rating of Understanding (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Understanding' for yoghurt.

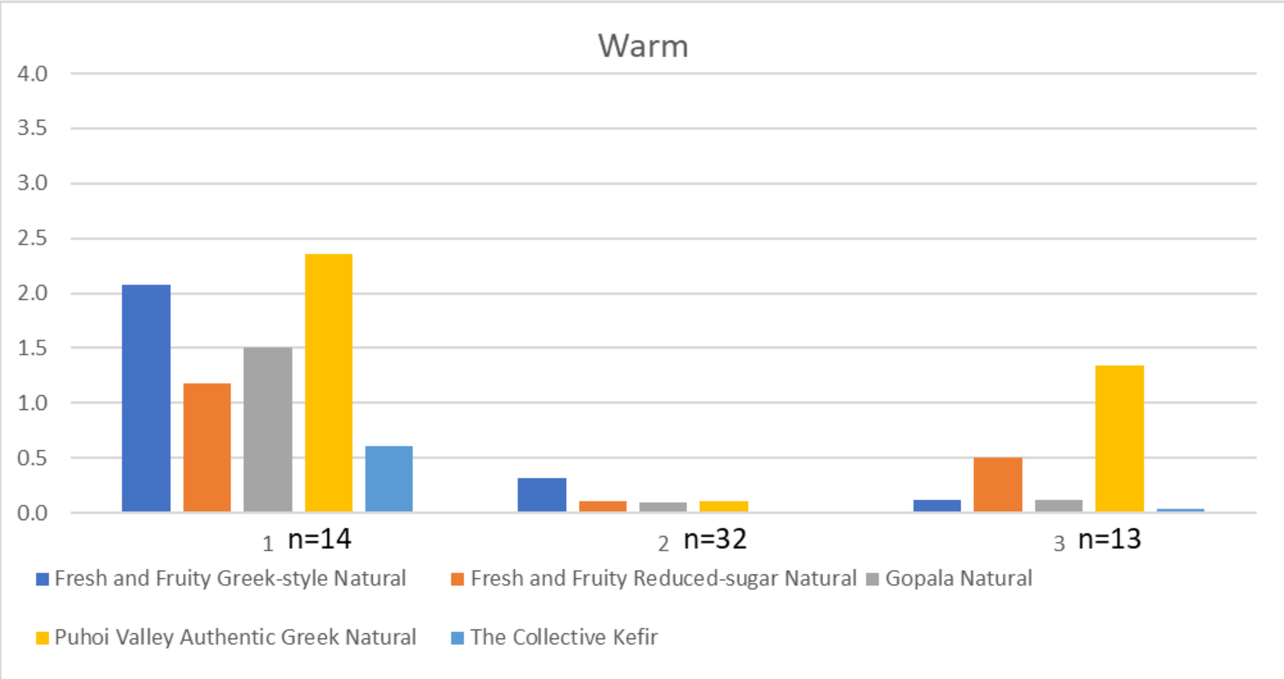

Figure A45: Mean rating of Warm (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Warm' for yoghurt.

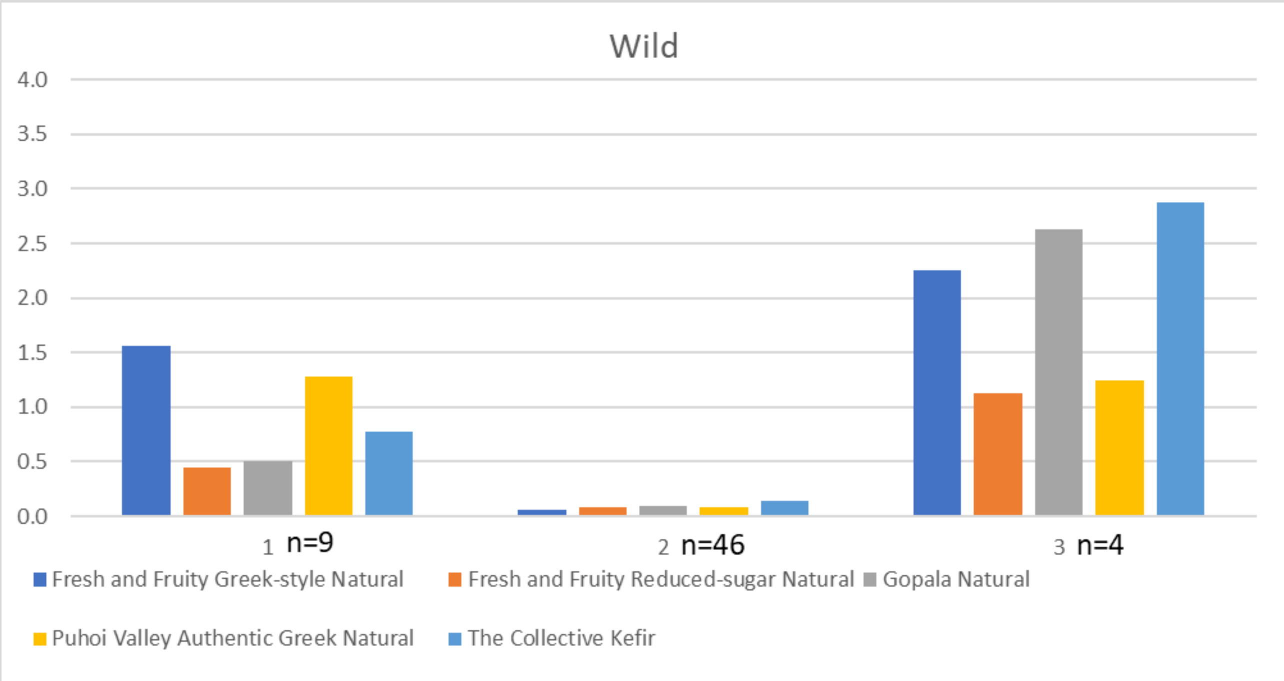

Figure A46: Mean rating of Wild (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Wild' for yoghurt.

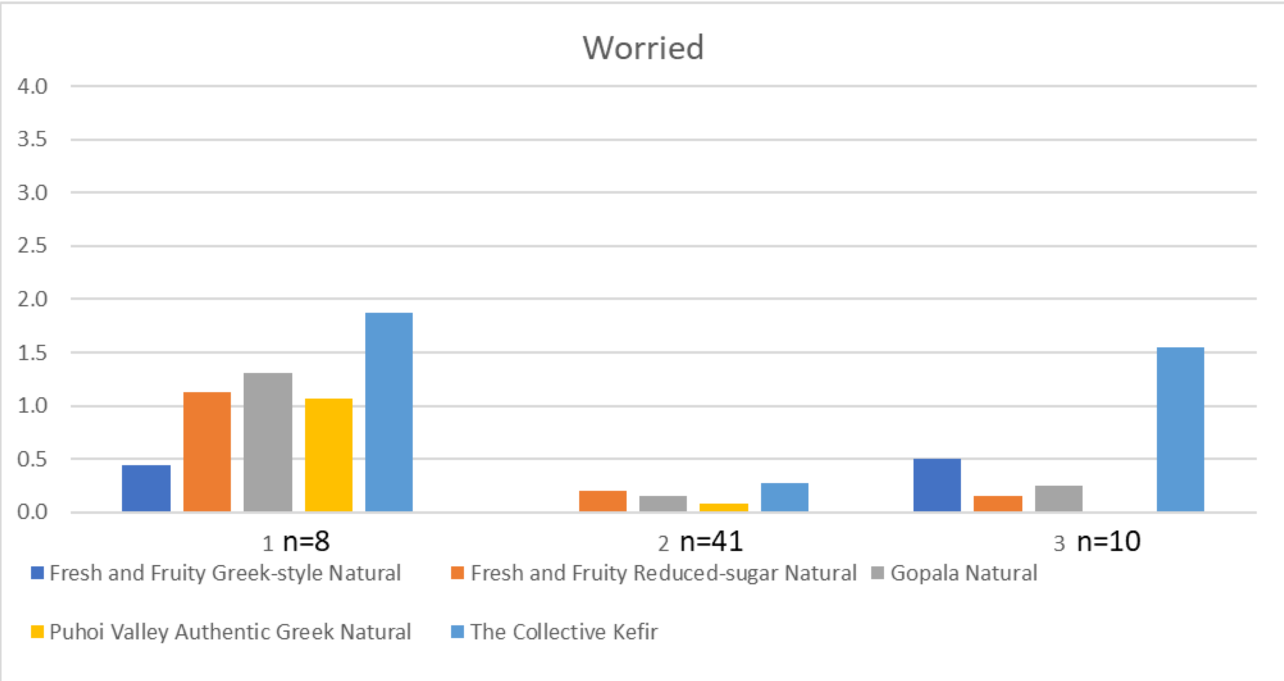

Figure A47: Mean rating of Worried (on a 0-4 scale) for each yoghurt product for the three clusters of participants grouped by their ratings of 'Worried' for yoghurt.