Methods to assess food-evoked emotions across cultures

Daisuke Kaneko

Methods to assess food-evoked emotions across cultures

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Wednesday 20th of April 2022 at 16.45 hours

by

Daisuke Kaneko

born on the 2nd of October, 1978
in Hakodate, Japan

This dissertation has been approved by:

Supervisor
prof. dr. J.B.F. van Erp

Co-supervisor
dr. A.M. Brouwer

Cover design: Daisuke Kaneko
Printed by: Ipskamp printing
ISBN: 978-90-365-5331-5
DOI: 10.3990/1.9789036553315

Graduation Committee:

Chair / secretary:	prof. dr. J.N. Kok (University of Twente)
Supervisor:	prof. dr. J.B.F. van Erp (University of Twente)
Co-supervisor:	dr. A.M. Brouwer (TNO)
Committee Members:	prof. dr. ir. H.J. Hermens (University of Twente)
	prof. dr. D.K.J. Heylen (University of Twente)
	prof. dr. ir. C. de Graaf (Wageningen University and Research)
	prof. dr. S.H. Fairclough (Liverpool John Moores University)
	dr. G. Jager (Wageningen University and Research)
	dr. S.D. Delplanque (Swiss Center for Affective Sciences, University of Geneva)

Contents

Chapter 1

General Introduction

In this introductory chapter the background and aim of the thesis are described in subsections on i) the definitions of emotions, ii) the link between food and emotions, iii) the current tools to evaluate food-evoked emotion, and iv) the importance of evaluating food-evoked emotion taking cross-cultural aspects into account. The chapter ends with a thesis outline.

What is an emotion?

During the 20th century, models of emotions have been developed within traditions of experimental psychology and neuroscience, and since the 1980s, accumulated experimental data have shown that emotion is relevant for attention, memory, decision-making, and action (Coppin and Sander, 2016). Emotion has been defined in various ways. One study (Kleinginna and Kleinginna, 1981) compiled and reviewed 92 definitions of emotion and 9 skeptical statements of emotion that oppose the usefulness of the concept "emotion" from the literature. Then, they classified those definitions and statements into 11 categories on the basis of the emotional phenomena or theoretical issues emphasized (see table 1-1). Since then, for instance, King and Meiselman (2010) define emotions as brief, intense physiological and mental reactions, Gibson (2006) defines emotions as short-term affective responses to the appraisal of particular stimuli, Bagozzi et al. (1999) define emotion as a mental state of readiness that arises from cognitive appraisals of events or thoughts, and Cabanac (2002) proposed that an emotion is any mental experience with high intensity and high hedonic content (pleasure/displeasure). Mehrabian and Russell (1974a) proposed Pleasure-Arousal-Dominance (PAD) emotional theory, which is a model on three different continuums: valence, arousal, and dominance. Sander (2013) suggested a consensual definition that “an emotion is an event-focused, two-step, fast process consisting of (1) relevance-based emotion elicitation mechanisms that (2) shape multiple emotional responses (i.e. action tendency, automatic reaction, expression, and feeling)”. These different definitions contain elements of both internal and external, bottom-up and top-down, as well as physiological and cognitive elements.

Table 1-1. Eleven categories of emotions on the basis of the emotional phenomena or theoretical issues emphasized.

Categories	Definitions	Basis
Affective Definitions	Emphasizing feelings of arousal and/or hedonic value	Subjective aspects
Cognitive Definitions	Emphasizing appraisal and/or labelling processes	
External Stimuli Definitions	Emphasizing external emotion-generating stimuli	Stimulus-Organism-Response (SOR) paradigm
Physiological Definitions	Emphasizing internal physical mechanisms of emotion	
Emotional/Expressive Behavior Definitions	Emphasizing externally observable emotional responses	

Disruptive Definitions	Emphasizing disorganizing or dysfunctional effects of emotion	Functional consequences
Adaptive Definitions	Emphasizing organizing or functional effects of emotion	
Multiaspect Definitions	Emphasizing several interrelated components of emotion	Scopes
Restrictive Definitions	Distinguishing emotion from other psychological processes	
Motivational Definitions	Emphasizing the relationship between emotion and motivation	
Skeptical Statements	Questioning or denying the usefulness of the concept of emotion	

It is clear that emotions can be elicited by a wide range of perceptual and symbolic stimuli (Bradley and Lang, 1994), including food stimuli. According to Russell (1980a), emotions can be described along the principal dimentions of valence (unpleasant and pleasant) and arousal (deactivation and activation), and the circumplex model of human core affect was proposed that orders emotions in a two dimensional space. Woodward et al. (2017) suggest that both valence and arousal play a distinct and critical role in eating-related behavior, and in a recent review paper, Prescott (2017) argues that the measurements of valence and arousal are more scientifically reliable in sensory and consumer food research than distinct types of emotions. In accordance with these recent arguments, in this thesis, emotions are regarded as the two principal dimentions of valence and arousal based on the cirumplex model, and the link beween food and emotion is described in the next section.

Link between food and emotion.

How is emotion elicited by food experience?

Sensing and interacting with food can evoke emotions. People respond to the sight of foods, the odor of foods, the flavor of foods, and even to the story of how a certain food is prepared. Also, elicited emotion by food experience is depending on emotional context (Kuenzel et al., 2011) and cultural background (Meiselman, 2015a). The adapted mechanism of how emotion is elicited by a certain food stimulus as described by Coppin and Sander (2016), is represented in Fig. 1-1. Appraisal processes refer to the subjective evaluation of food products’ intrinsic cues (e.g. sensory perception of food itself) and extrinsic cues (e.g. contexts in which food is presented). Embodied states refer to the activation of various bodily-related systems (e.g. individual’s food habit and habitual attitude to food). Core relational themes represent more categorical conceptualization of emotion-eliciting appraisals, such as and individual’s conceptualization of food based on cultural backgrounds and food trait based on traditional trend in one region. Core affect is a consciously accessible emotional state resulting from a combination of two continuums (the circumplex model of emotion): valence (pleasant to unpleasant) and arousal (low to high activation) (Russell, 1980a; Russell and Barrett, 1999). Valence is an important determinant in food liking but may be considered of limited value in consumer research when measured on its own, as most products elicit equally high liking scores within their food category (He et al., 2016a; Beyts et al., 2017). Arousal has received relatively less attention in emotional food research, but this may be undeserved. Arousal affects the memorability of an event where an adequate level of arousal is required for a product to be remembered and eaten again in the future (Köster and Mojet, 2007).

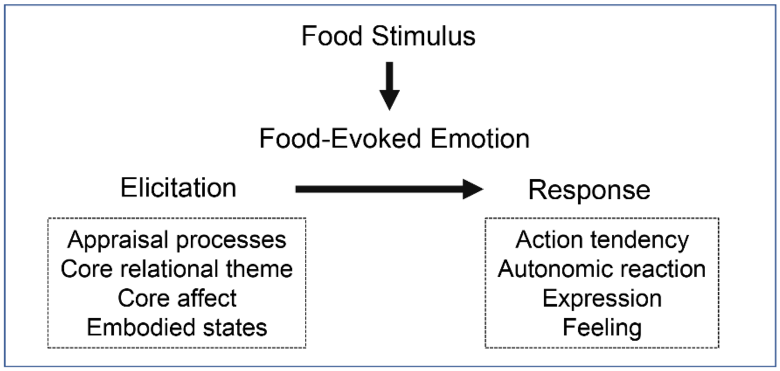


Figure 1-1. Mechanisms of food-evoked emotions and their effects on emotional response (adapted from Coppin and Sander (2016))

There are multiple responses to emotion elicitation, including action tendency, autonomic reaction, expression, and feeling. In the food discipline, action tendency usually refers to tendencies of approach (the tendency to move toward stimuli that are intuitively appraised as good or beneficial) and avoidance (the tendency to move away from stimuli that are intuitively appraised as bad or harmful) (Kemps et al., 2013; Piqueras-Fiszman et al., 2014; Becker et al., 2015; Brockmeyer et al., 2015). Autonomic reaction refers to the response of the Autonomous Nervous System (ANS), e.g. changes in heart rate. Expression is also considered as one of emotional responses to food, a clear example of expressing emotional responses to foods (e.g. disgust, comfort, pleasantness, etc.) is facial expression. Many studies use “feeling” and “emotion” interchangeably in everyday language (Coppin and Sander, 2016). In this thesis, “feeling” is considered as being inclusive of “emotion”. Also, in a recent review, de Wijk and Noldus (2021) describe appraisal theories emotions in the food domain such as the component process model proposed by Coppin and Sander (2016).

Why is it important to evaluate food-evoked emotions?

There is a reciprocal relation: food experience affects one’s emotional state, and emotional state affects one’s food behavior such as food choice and eating behavior (Bellisle, 2009). The study by Richins (1997) first evaluated the emotions evoked by consuming foods. A series of studies by Macht and colleagues (Macht (1999); Macht and Simons (2000); Macht et al. (2002)) show that there are two “emotion congruent effects” between emotion and eating behavior. First, negative emotions, such as “anger”, “sadness”, and “fear”, evoke more eating behavior. Second, similar to negative emotional state, positive emotional state, such as “pleasantness” also induce people to eat more. Compared to negative emotional state, they take more time to enjoy tasty foods. More recent studies also demonstrated that emotions can be determinants of food-related behavior including food choice and eating behavior (e.g. (Russell, 2003; Gibson, 2006; Bellisle, 2009; King and Meiselman, 2010; Thomson et al., 2010; Ng et al., 2013; Dalenberg et al., 2014; Gutjar et al., 2015a; Kenney and Adhikari, 2016b)). Gutjar et al. (2015b) suggested that food-evoked emotion is a more important predictor for an individual's food choice compared to liking. Also, Dalenberg et al. (2014) mentioned that consumers' emotions add predictive power to a food choice (predicting) model based on hedonic scales. These studies suggest that assessing emotional responses to foods may reveal previously unknown product attributes which can be a valuable source of information for product development and marketing that goes beyond traditional sensory and acceptability measurements (Thomson et al., 2010). This leads to the general question what type of measures (and what combinations of measures) are valid, reliable and efficient to capture one's emotional responses to food, and how they are practical under cross-cultural condition.

Tools to measure food-evoked emotions.

Explicit and implicit measures to evaluate food-evoked emotions.

In this thesis, we define three levels of emotional processing: the lower level (the unconscious and basic sensory processing), the intermediate level (perception and early cognitive processing), and the higher level (the conscious processing stage after cognition including decision-making and food related behaviour) adapted from a conceptual multisensory response model provided by Schreuder et al. (2016). The term "explicit" is linked to the conscious processing, and "explicit measure" refers to conscious, cognitive responses to foods, such as self-report ratings. On the other hand, the term "implicit" denotes the unconscious and early cognitive processing, and "implicit measure" refers to a measure that captures the unconscious physiological and behavioural responses, such as heart rate and electrodermal activity, and facial expression and approach-avoidance tendency, respectively.

In the past years, a number of explicit emotion questionnaires have been developed for clinical purposes and academic research, such as The Profile of Mood States (POMs; (McNair et al., 1971)) and The Positive and Negative Affect Schedule (PANAS; (Watson et al., 1988)). These questionnaires have not been specifically developed to measure food-evoked emotions and emphasize negative emotions. However, given that food items usually elicit positive rather than negative emotions, positive emotions can be argued to be more relevant in selecting and consuming foods. Therefore, it is also necessary to have a measurement tool that is able to capture more positive emotions. The study by Richins (1997) is one of the first that measured food and product emotions and proposed consumption emotion sets consisting of 47 emotional words constructed to gather emotional information toward "consumption" experiences including foods. This consumption emotion set was a starting point and triggered the development of several different measurement tools to assess food-evoked emotions. One of the most widely used explicit self-report questionnaires to measure food-evoked emotion is the EsSense Profile (King and

Meiselman, 2010), and a shorter version called EsSense25 (Nestrud et al., 2016b) that is easier and more cost-effective to apply while retaining discriminative power (Churchill and Behan, 2010; Dorado et al., 2016). There is still serious debate on what the best method to evaluate food-evoked emotion is within enormous explicit self-report measures developed at the present (Meiselman, 2021). However, while the explicit self-reports are the most popular measure to evaluate food-evoked emotions, there are inherent drawbacks: they are discontinuous (they reflect a single moment in time), may suffer from response biases, and may cover only cognitive processes (Venkatraman et al., 2015). In addition, emotional words vary across cultures and languages, and individuals are not always used to verbalizing their emotions in their daily life, particularly for food-evoked emotions (Gutjar et al., 2015b). Finally, Winkielman et al. (2011b) argues that when individuals are instructed to explicitly rate their emotions, this may interfere their actual food experience itself.

Besides the explicit self-report measures, more recent studies introduce implicit measures such as behavioral and (neuro) physiological measures to capture unconscious emotional response to foods. Examples are ANS responses to odors; (de Wijk et al., 2012; De Wijk and Boesveldt, 2016), facial expression to a taste of a drink; (de Wijk et al., 2014; Kostyra et al., 2016), and approach-avoidance motivation to foods; (Piqueras-Fiszman et al., 2014). As mentioned, advantages to use these implicit measures would be that they are able to evaluate continuous change of food-evoked emotions before, during, and after food experience while explicit self-report emotional questionnaires measure emotion at a single moment in time. On the other hand, it is difficult to estimate which emotion people experience with only one single implicit physiological measure. In accordance with this dilemma, it is plausible that there is not a single measure that would be able to capture the full range of relevant emotional responses to foods. Thus, it seems worth the effort to combine explicit and implicit types of measurements in order to compare and eventually cross-validate results (Köster, 2009). Attempting to measure all emotional processing levels of food-evoked emotions at the same time would be expected to bring important insights in the food domain. A review by Bell et al. (2018) also argues that combining explicit and implicit measures may complement the research data to better predict food-evoked emotions. However, to the best to our knowledge, there is no ‘golden standard’ measure or a combination of measures to assess food-evoked emotions yet, which is more elaborately elucidated in the literature review in Chapter 2.

Methodological issues: ground truth, validation aspect

In studies in experimental psychology and human factors, it is important to define emotions or cognitive constructs that are of the study’s interest and to clearly define the ground truth used in the study (Brouwer et al., 2015c). In emotion research in the food domain, it may be no exception, especially when a study uses implicit (neuro)physiological measures to evaluate food-evoked emotions. One way to solve this requirement is to explicitly ask participants to rate their emotions evoked by food experience and consider that as the ground truth. However, subjective ratings do not necessarily match with physiological responses (Brouwer et al., 2015c) as they measure different aspects of the emotional process as argued above. In food sensory studies, expert judgement of taste and flavor by trained panels can be set as a ground truth of a certain taste of food, but for ground truth of emotion this would not make sense. Although there have been several studies investigating food-evoked emotions with implicit (neuro)physiological measures, it is currently too early to provide (a set of) implicit measures as a ground truth, and few studies set a ground truth of food-evoked emotions without explicit ratings. In order to justify how emotional responses to food experience are related to the results of implicit behavioral and physiological measures, one can use a stimulus that is prior known to evoke similar emotional responses across participants. In Chapter 4, we describe this approach and use two different ways of setting the ground truth to train a machine learning model to evaluate to what extent the model is able to predict participants' emotions (valence and arousal) evoked by food experience without explicit self-report ratings.

Importance of when to evaluate emotions: during and after food experience.
The majority of studies collect participants' emotions evoked after the food experience and focus on the instantaneous effect only. Although emotions are considered as short-term, intense responses to food experience (Meiselman, 2015a), Köster and Mojet (2015b) argue that the role of memory is probably much more important than the "first impression" experience. They emphasize that food-evoked emotions should be tested before, during, a few hours after, and a week (or even longer) after the food experience, to obtain a more complete picture of the experience of the product. This is more elaborately discussed in Chapter 5.

Significance of measuring food-evoked emotions across cultures.

Emotional expression is not the same between East and West (between American and East Asian (Tsai, 2007), between Europe and Iran (Riegel et al., 2017)), and even among different Western countries (Meiselman, 2015a), such as between European American and Hmong American (Tsai et al., 2002) and between different Spanish speaking countries (van Zyl and Meiselman, 2016a). Cross-cultural studies on food-related emotions are becoming increasingly important, especially for large food companies that have global markets as a result of the globalization of food products (Meiselman, 2013b; Meiselman, 2015a).

However, verbal self-assessment tools typically pose difficulties for cross-cultural research since emotional words are often not directly equivalent in different languages (Wierzbicka, 1986; 1999). In addition, consumers from different cultures tend to use emotional terms differently (Meiselman, 2015a; van Zyl and Meiselman, 2015; van Zyl and Meiselman, 2016a). Researchers use back-translation methods to try and avoid the pitfalls of subtle differences in meaning among emotional lexicons between different countries. However, it is argued by Boster (2005a) that there may be some misunderstandings among cultures in emotion intensity, context and other semantics when using backtranslation approaches for the translation of emotion words in different countries. Also, consumers' response styles are different depending on ethnic groups. For example, Western respondents typically show an extreme response style (ERS), whereas Asian respondents tend to be reluctant to strongly express negative opinions and as a consequence seem to typically show a middle response style (MRS) (Chen et al., 1995; Lee et al., 2002; Lottridge et al., 2011), which may cause a misinterpretation of their actual emotions. Besides the conventional measures of verbal self-assessment, non-verbal explicit and implicit measures may therefore be valuable tools in cross-cultural studies since they are independent of language. This thesis also aims to investigate if these non-verbal explicit and implicit measures are sensitive enough to evaluate emotions evoked by the same food in consumers with different food cultural backgrounds, which is mainly discussed in the second part of the thesis.

Thesis outline

This thesis consists of two main parts. The first part, 'Measures to evaluate food-evoked emotion', i) discusses what (type of) measures are used to evaluate food-evoked emotions, ii) investigates to what extent present measures are able to capture food-evoked emotions, iii) examines if combining measures in a model improves the sensitivity, and iv) finally evaluates the effect of emotional context on the pleasantness of novel and familiar foods with both explicit and implicit measures. These research topics are presented in Chapter 2, 3, 4, and 5, respectively.

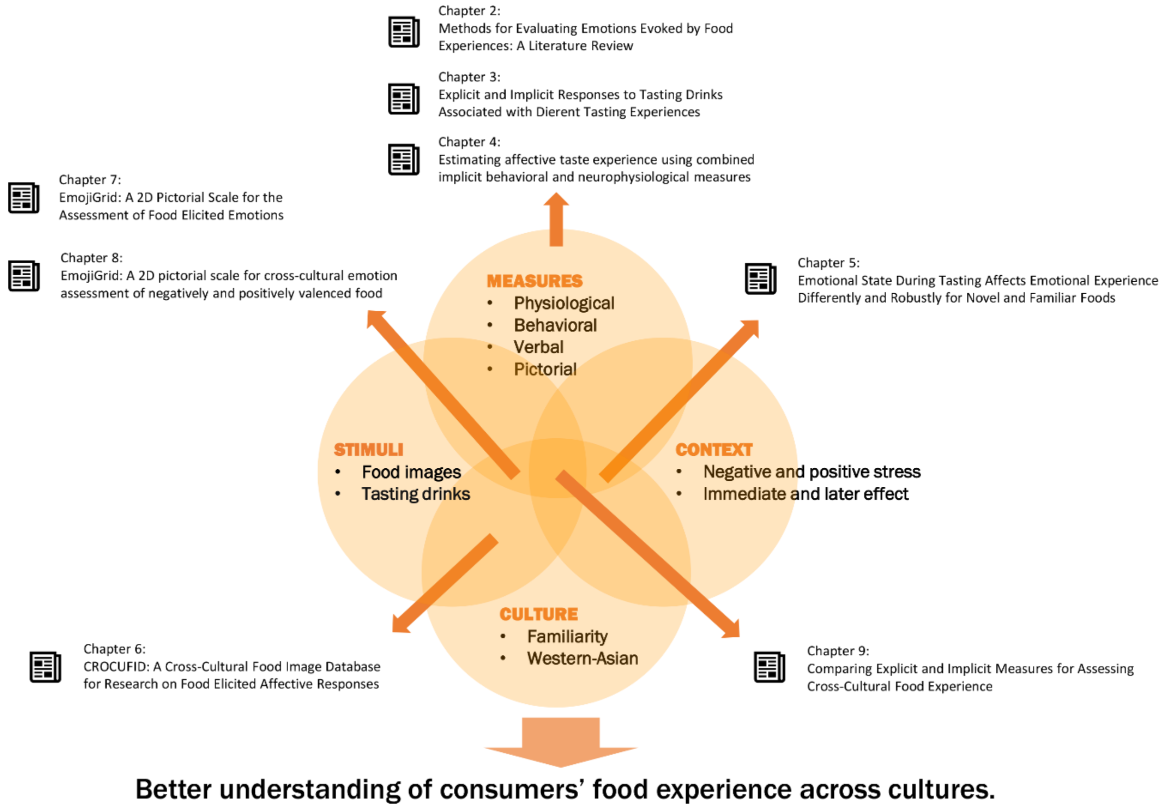


Figure 1-2. The schematic overview of this thesis (Knowledge in food experience)

The second part focuses on 'Measures of food-evoked emotion in a cross-cultural context'. In Chapter 6, a new food image database including not only Western food menu but also Asian dishes is introduced and validated for cross-cultural studies. Chapter 7 validates a newly developed language-independent tool to evaluate food-evoked emotion across cultures. Chapter 8 investigates to what extent the new tool reliably assesses food-evoked emotions by examining responses to the same universal food images of consumers from different cultures. Lastly, a cross-cultural study is conducted in the Netherlands and Thailand to investigate the hypothesized advantages and disadvantages of explicit and implicit measures in practice, discussed in Chapter 9. A schematic overview of the thesis is given in figure 1-2.

Part I - Measures to evaluate food-evoked emotion.

In Chapter 2, we present an exhaustive list of measurement tools that have been used to measure food-evoked emotions, extracted from the literature over the last 20 years. Measures are classified as explicit (e.g. self-report questionnaires) and implicit (e.g. (neuro) physiological measures). In previous consumer related studies, explicit, verbal measures have been used most often. However, these may be biased e.g. due to demand characteristics. Implicit physiological responses may therefore have a potential to complement explicit measures to understand one's food-evoked emotions. However, implicit measures may not be specific and sensitive enough, and it is more difficult to link them to the ground truth. We present a 'look-up-table' as guidance to combine explicit and implicit measures to evaluate food-evoked emotions as a function of emotional processing level (unconscious, perceptual, and cognitive levels).

The literature review indicates that implicit methods could be beneficial to complement the subjective ratings. Therefore, in Chapter 3, we compared a range of explicit and implicit methods to measure food-evoked emotions with the following research questions: (i) how well do different self-report, physiological, and behavioral variables discriminate regular drinks from a drink that is known to be strongly disliked (vinegar)? (ii) how sensitive are these measures to the subtle differences between regular drinks? (iii) how are different implicit measures associated with self-reported valence and arousal for regular drinks? In Chapter 4 a linear model using combinations of implicit variables was trained to distinguish between vinegar (a ground truth unpleasant) and regular drinks, and applied to data of regular drinks to investigate to what extent the trained model could detect relatively subtle differences between emotions elicited by regular drinks.

In Chapter 5, we further applied multiple explicit and implicit measures to investigate the effect of affective context in which food is experienced on the emotional response to novel and familiar foods both during initial tasting and a week later, addressing the reciprocity between emotional state and food and the recommendation by Köster (2009) to not only probe food-evoked emotion immediately after food experience.

Part II - Measures of food-evoked emotion in cross-cultural context.
In Chapter 6 to 9, we study food-evoked emotion using multiple explicit and implicit physiological measures across cultures.

In order to conduct cross-cultural studies, it is often necessary to use food stimuli that represent food from different cultures. Since these were not publicly available, we created a publicly available and extendable food image database called CROCUFID (a CROss-CUltural Food Image Database) with high-resolution, standardized images from various (currently mainly Western and Asian) cuisines. In Chapter 6, we presented and validated the CROCUFID.

While Chapter 3-5 show the sensitivity of physiological measures, and while physiological measures have certain advantages, the advantage of questionnaires is their ease of application and face validity. In Chapter 7, we design and validate a new affective self-report tool called "EmojiGrid" that has these advantages, but does not have the common disadvantage of questionnaires: misunderstanding derived from language translation. Emotion related language is hard to translate into different languages, and a large variability in interpretation may exist, as described in section I-4.

In order to evaluate if the EmojiGrid goes across cultures, we conducted a cross-cultural study with EmojiGrid as a rating tool in Chapter 8. In this study, we investigated how individuals with different mother tongues (English from UK, Dutch from the Netherlands, German from Germany, and Japanese from Japan) respond to the same set of 60 universal food images with EmojiGrid.
Finally, in Chapter 9, we combine all types of measures in a study performed in the Netherlands and in Thailand in which participants view and taste culturally familiar and unfamiliar food. We provide direct evidence that some of the implicit measures can overcome culturally-related biases observed in explicit measures.

Chapter 10 provides the integrated discussion and general conclusion of results obtained in this thesis.

Chapter 2

Methods for Evaluating Emotions Evoked by Food Experiences: A Literature Review

This chapter is published as:
Kaneko, D., Toet, A., Brouwer, A.M., Kallen, V., and Van Erp, J.B. (2018). Methods for evaluating emotions evoked by food experiences: a literature review. Frontiers in psychology 9, 911.

Abstract

Besides sensory characteristics of food, food-evoked emotion is a crucial factor in predicting consumer's food preference and therefore in developing new products. Many measures have been developed to assess food-evoked emotions. The aim of this literature review is (i) to give an exhaustive overview of measures used in current research and (ii) to categorize these methods along measurement level (physiological, behavioral, and cognitive) and emotional processing level (unconscious sensory, perceptual / early cognitive, and conscious / decision making) level. This 3x3 categorization may help researchers to compile a set of complementary measures ('toolbox') for their studies. We included 101 peer-reviewed articles that evaluate consumer's emotions and were published between 1997 to 2016, providing us with 59 different measures. More than 60% of these measures are based on self-reported, subjective ratings and questionnaires (cognitive measurement level) and assess the conscious / decision-making level of emotional processing. This multitude of measures and their overrepresentation in a single category hinders the comparison of results across studies and building a complete multi-faceted picture of food-evoked emotions. We recommend (1) to use widely applied, validated measures only, (2) to refrain from using (highly correlated) measures from the same category but use measures from different categories instead, preferably covering all three emotional processing levels, and (3) to acquire and share simultaneously collected physiological, behavioral, and cognitive datasets to improve the predictive power of food choice and other models.

Introduction

People experience and appreciate many types of food and beverages (referred to as ‘foods’ in this study) during their life. Specific emotions have been considered as determinants of affective responses to foods (Ferber and Cabanac, 1987; Willner and Healy, 1994) and food-related behavior including food choice (e.g. (Oliver and Wardle, 1999; Gibson, 2006); (Russell, 2003; Thomson et al., 2010; Ng et al., 2013; Dalenberg et al., 2014; Gutjar et al., 2015a; Kenney and Adhikari, 2016b)). Liking ratings do not predict food choice behavior accurately (Zandstra and El-Deredy, 2011; Griffioen-Roose et al., 2013a). Gutjar et al. (2015b) suggested that food-evoked emotions can predict individual’s food choice more accurate than liking scores. Dalenberg et al. (2014) mention that consumers’ emotions add predictive power to a food choice (predicting) model based on hedonic scales. These studies suggest that assessing emotional responses to foods may reveal previously unknown product attributes which can be a valuable source of information for product development and marketing that goes beyond traditional sensory and acceptability measurements (Thomson et al., 2010). Therefore, it is important to obtain valid and reliable (combinations of) measurements of food-evoked emotion.

Despite its importance, different authors use different definitions of emotion. For instance, King and Meiselman (2010) define emotions as brief, intense physiological and mental reactions, Gibson (2006) defines emotions as short-term affective responses to the appraisal of particular stimuli, Bagozzi et al. (1999) define emotion as a mental state of readiness that arises from cognitive appraisals of events or thoughts, and Cabanac (2002) proposed that an emotion is any mental experience with high intensity and high hedonic content (pleasure/displeasure). These different definitions contain elements of both internal and external, bottom-up and top-down, as well as physiological and cognitive elements. All these facets are considered relevant and illustrate that there is not a single measure that would be able to capture the full range of relevant aspects. To organize the complex response patterns, we introduce a conceptual framework in section 3, including the methodologies to assess the different response patterns as there is a wide variety of instruments available.

Verbal self-reporting questionnaires are the most commonly used techniques to measure emotional responses, due to their ease of application, cost-effectiveness and discriminative power (Dorado et al., 2016); (Churchill and Behan, 2010). However, they have specific shortcomings, including: (1) emotions are difficult to verbalize (Köster and Mojet, 2015b), (2) the ‘emotional’ lexicon varies across cultures and languages, particularly when it comes to foods (Gutjar et al., 2015b), (3) verbalizing emotions can interfere with the food experience itself, and (4) self-reports only capture conscious, declared opinions (Venkatraman et al., 2015);(Winkielman et al., 2011b). Wilson et al. (1993) asked their participants to answer whether they liked or disliked five different posters with or without providing the reason why. Subsequently, they could take one of the posters with them. Participants who provided the reasons were less satisfied with their choice three weeks later (Wilson et al., 1993) showing that questioning individuals about affective experience can affect the affective experience itself. Regarding EsSense Profile, one of the most widely used self-report questionnaires for evaluating an individual's emotional responses (King and Meiselman, 2010), Jaeger et al. (2013) stated that this technique might not capture the full range of emotions individuals may experience in response to food and therefore may not properly measure food-evoked emotions. Thus, it seems worth the effort to include other types of measurements as well like behavioral and physiological measurements. Köster (2009) proposed that research groups should develop implicit measurement techniques and use these where possible and combine them with explicit measures if feasible in order to compare and eventually cross-validate results Examples include facial expression recognition (happy, sad, angry, surprised, scared, and disgust: (Kostyra et al., 2016) and physiological variables reflecting activity of the autonomic nervous system (ANS: (de Wijk et al., 2012)). However, there is no 'golden standard' to assess food-evoked emotions at this moment yet.

We aim to provide an exhaustive list of tools that have been used to measure food-evoked emotions over the last 20 years. We also categorize them using a general model describing the relevant aspects of emotion processing and the range of methodologies to assess the relevant facets (see section 3). This categorization helps to identify gaps in the currently prevailing set of instruments and enables researchers to choose (a combination of) measures in a balanced way. Our categorization indicates to what extent different methods are redundant or complementary and helps researchers in this area to compile a set of complementary methods that provides the maximal amount of information. In addition, it may serve to guide further development of new methods to assess food evoked emotions that predict future consumer behavior.

Literature search

We used the databases of PsycINFO to select relevant articles that were published between January 1997 until the end of December 2016.

Inclusion criteria

We used the following inclusion criteria:

1. The article should report empirical studies in peer-reviewed journals and be written in English.
2. The article should include original data from healthy human populations.
3. The study should investigate consumers' emotions evoked by directly experiencing foods. 'Direct experience' could be tasting foods, viewing images of food, or sniffing food odorants. The following were considered to be indirect (and therefore excluded from this review): viewing packages, viewing printed names of brands and thinking about food or beverages (e.g., by asking "*How do you feel when you think of 'apple'?*").
4. In this study the term 'emotions' includes hedonic liking, pleasantness, preference, and moods. Gibson (2006) describes moods as more long-lasting psychological arousal states than emotions with interacting dimensions related to energy, tension and pleasure that may appear and persist in the absence of obvious stimuli and may be more covert to observers. However, mood and emotion both reflect emotional states and are often used interchangeably in common language (Köster and Mojet, 2015b). Also, as Gibson (2006) mentions, the relationships between mood, emotions and

physiological arousal may be complex. Therefore, we included ‘mood’ in our criteria. While the sensory characteristics of food (e.g., appearance, aroma, taste and texture) are important drivers of emotional experience, we here focus on methods to measure food-evoked emotion, and not on methods to assess the perception of sensory characteristics (e.g., this tastes sweeter and feels softer than others). Thus, studies on the appraisal of the qualitative characteristics of foods (e.g., intensity of sweetness, sourness, saltiness, spiciness, and bitterness) are not included here.

Search procedure

Three reviewers (DK, AT, and AB) constructed the inclusion criteria, searched and evaluated the relevant literature. To obtain relevant articles from the PsycINFO database, the following combination of keywords was used: (food *OR* foods *OR* beverage *OR* beverages) *AND* ((‘explode’ emotions *OR* emotional responses *OR* emotional states *OR* physiological arousal) *OR* (pleasantness *OR* hedonic *OR* liking *OR* preference)). We used the ‘explode’ function in the PsycINFO search tool. For instance, exploding ‘emotions’ provides all articles related to emotions. In addition to exploding emotion-related keywords, we further searched relevant articles using keywords such as pleasantness, hedonic, liking, and preference. As a result, we obtained an initial pool of 9,873 articles. Then, by limiting our search to articles in English reporting empirical studies in peer-reviewed journals, 8,156 articles were selected. Among them, we excluded articles targeting animals and disordered populations and kept 3,031 articles. We finally obtained 2,355 articles published in a 20 year period ranging from 1997 to 2016. Based on reviewing the title and abstract of those articles, we excluded articles with a lack of relevance (i.e., they did not meet our inclusion criteria as described above) and ended up with 65 articles. Then, full-text screening resulted in 57 relevant articles. For most of the relevant papers, measuring food-evoked emotions was not the main topic but part of the methodology to answer a different question of interest. This is why it proved difficult to capture *all* relevant articles using keywords like the ones listed above. Another 44 relevant articles were extracted based on cited references in the set of 57 articles and based on searches for more work by the first author, resulting in a final set of 101 papers. The eligibility of these additional 44 articles was independently assessed and confirmed by all three reviewers via an in-depth critical full-text review. A schematic representation of the search procedure is shown in Figure 2-1.

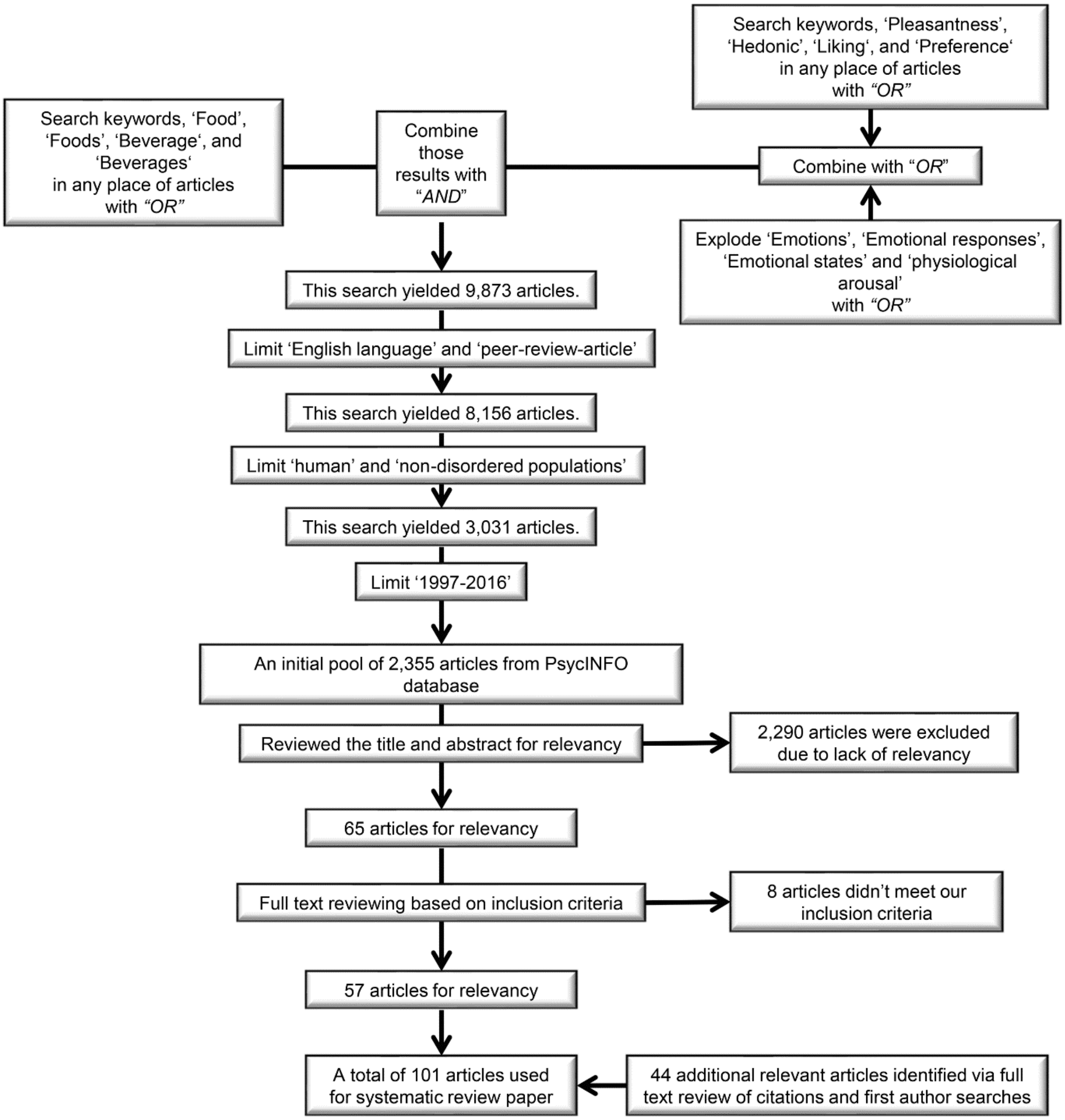


Figure 2-1. Schematic representation of literature search and selection procedure.

Overview of selected articles

Table S2 gives a summary of the final set of 101 articles about the stimuli, the methods to measure food-evoked emotions, and the key findings provided by those methods. More than half of these articles were published in the last four years, and about ten times more articles were published in the last four years than during the first four years (Figure 2-2). This suggests that there is a growing interest in understanding emotions evoked by foods. Figure 2-3 gives an overview of the stimuli that have been used to evaluate food-related emotions. Actual foods were used by far most often as stimuli. Representative foods stimuli were sweet products, such as chocolates and cakes, while savory foods were less frequently selected as stimuli. Most studies evaluated an individual's emotions for a sole product, not for a full meal. The vast majority of measures were conducted just before, during and right after experiencing foods, although some studies asked participants to report their emotions a certain amount of time after experiencing the sample stimuli.

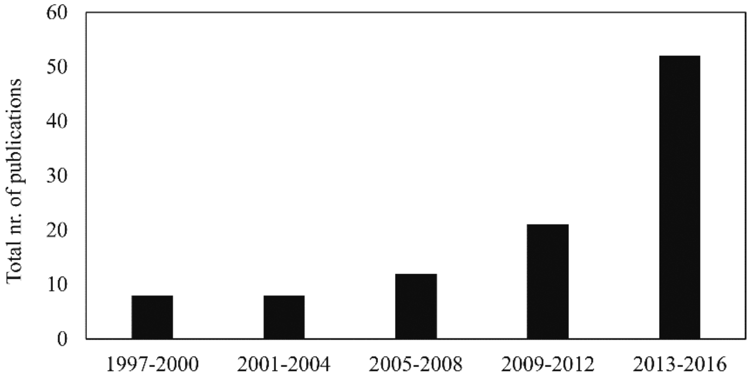


Figure 2-2. Total number of publications over successive 4-year intervals from 1997 to 2016.

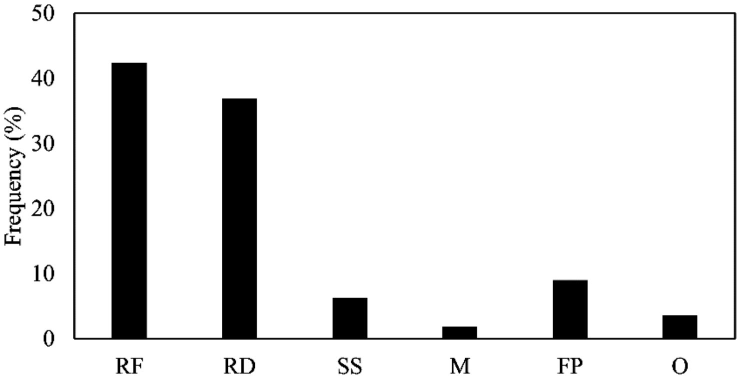


Figure 2-3. The frequency of stimuli used within 101 studies (RF, Regular solid foods; RD, Regular drinks; SS, Simple Solutions; M, Meal; FP, Food Pictures; O, Odors).

From the 101 selected articles, we identified a total of 59 different measures for the assessment of food-evoked emotions. Table 2-1 presents a brief description of each of these 59 different measures and a reference to a more elaborate description. The total number of times each measure was used within the selected group of 101 relevant studies is depicted in Figure 2-4. In this figure, we also grouped the measures according to general type of methodology (physiological, behavioral and cognitive). More than 80% of the papers used hedonic scaling measures to evaluate food-evoked emotions, indicating that this measure is most often used in addition to other measures. Following hedonic scales are several versions of emotional lexicon questionnaires such as CD-CATA, EsSense Profile, and unique instruments created by researchers themselves. Recording facial expression, usually by analyzing picture or movie data or electromyography (EMG), is a popular behavioral/implicit method. It should also be noted that more than 50% of measures were only used once among the 59 measures extracted from the 101 selected articles, indicating that there is still no representative measure or combination of measures developed for the evaluation of food-evoked emotions.

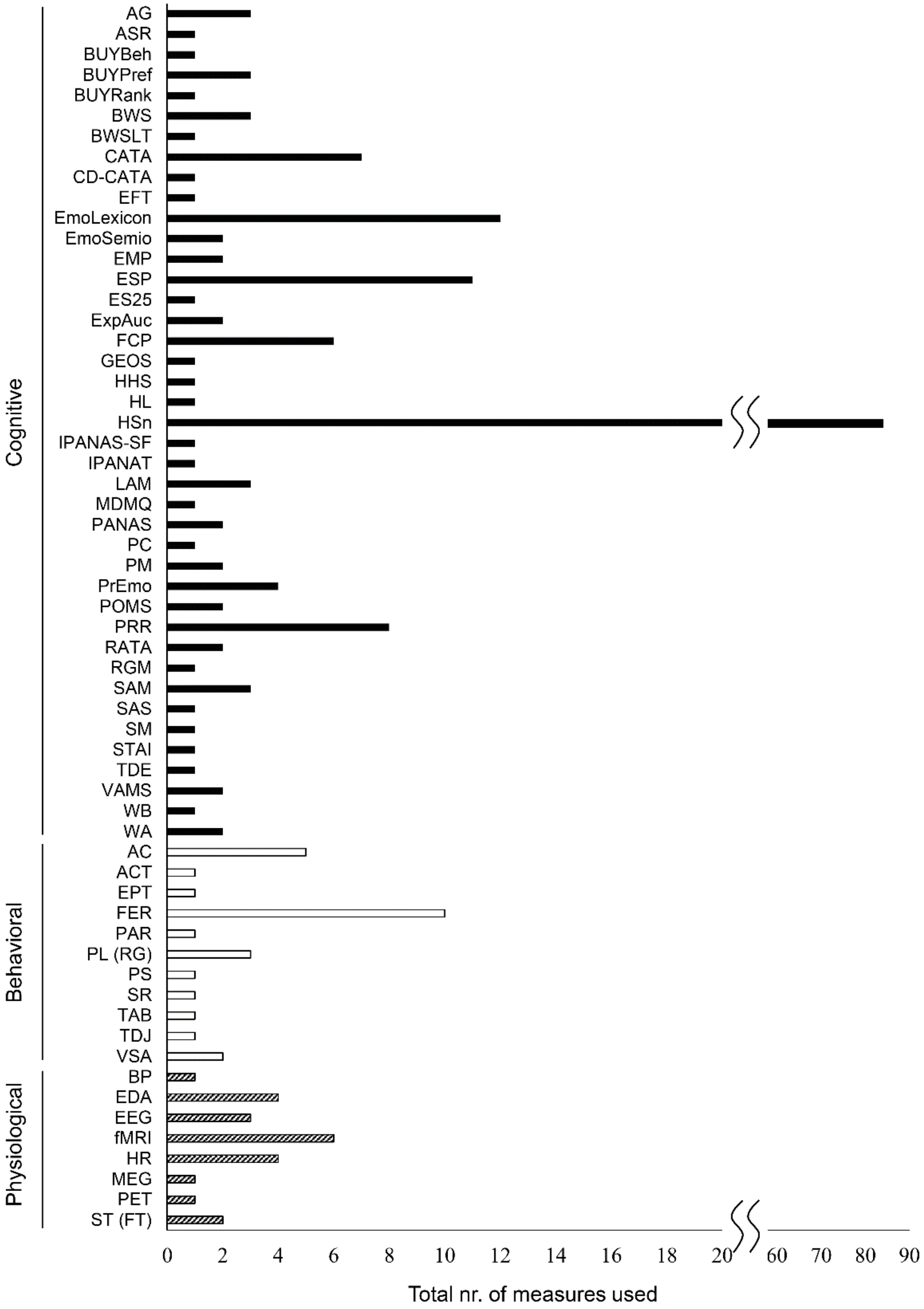


Figure 2-4. Total number of studies (from the selected set of 101) in which each of the measures is used. The black bars, the white bars, and the shaded bars represent the cognitive, behavioral, and physiological measures, respectively (All abbreviations are described in **Table 2-1**).

Table 2-1. Brief description and abbreviation of measures used for the assessment of food elicited emotions.

Method	Abbreviation	Description	References
Amount Consumed	AC	The weight, volume or number of food or drink products that are consumed. This measure tends to increase with hedonic evaluation.	Zandstra et al., 1999
Autobiographical Congruency Test	ACT	The ease and speed with which people remember sad or cheerful events in their lives are higher when they are congruent with their present emotional state.	Mojet et al., 2015
Affect Grid	AG	Two-dimensional scale to assess affect along the dimensions pleasure and arousal.	Russell et al., 1989
Affect Self Report scale	ASR	A self-report scale with 18 affective terms that can be scored on 7-point scales.	Christie and Friedman, 2004
Blood Pressure response	BP	Blood pressure is the pressure of circulating blood on the walls of blood vessels, usually expressed in terms of the systolic (maximum during one heart beat) pressure over diastolic (minimum in between two heart beats) pressure, and measured in millimeters of mercury (mmHg), above the surrounding atmospheric pressure. Blood pressure may change in response to changes in mood or emotions.	Bercea, 2013
Buying Behavior	BUYB	Actual buying behavior (buying frequency).	Rosas-Nexticapa et al., 2005
Buying Preference rating	BUYP	Self-reported likelihood to buy a product.	Rosas-Nexticapa et al., 2005
Buying Preference ranking	BUYR	Ranking different products according to self-estimated likelihood to buy.	Rosas-Nexticapa et al., 2005
Best-Worst Scaling or Maximum Difference Scaling	BWS	Assessors are presented a series of sample triads or tetrads from which they select the (best and worst) samples representing the largest difference in an underlying continuum, e.g. liking.	Jaeger et al., 2008
Best-Worst Scaling of Lexicon Terms	BWSLT	For different combinations of 5 words from a larger lexicon, choose which word most/least closely reflects product experience.	Thomson et al., 2010
Check-All-That-Apply	CATA	Assessors are presented with a list of sensory terms or phrases and are asked to select all those terms or phrases they consider applicable for describing the focal sample.	Adams et al., 2007

Table 2-1. Continued

Method	**Abbreviation**	**Description**	**References**
Consumer-Defined Check-All-That-Apply	CD-CATA	Assessors are presented with a list of 36 emotion terms elicited from interviews with consumers and are asked to select all those terms they consider applicable for describing the focal sample.	Ng et al., 2013
Electrodermal Activity	EDA	Electrodermal activity measures changes in the electrical resistance of the skin which reflects activation of the sweat gland in reaction to emotional stimuli. EDA is the generic term for all types of skin conductance variables such as skin conductance response (GSR) and skin conductance level (SCL).	Kreibig, 2010
Electroencephalography	EEG	An electrophysiological monitoring method to record electrical activity of the brain. It is typically noninvasive, using electrodes placed along the scalp. EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain.	Agarwal and Xavier, 2015, Bercea, 2013
Empathic Food Test	EFT	Assessors report their feelings after food consumption using a list of 12 empathic terms rated on 5-point scales.	Geier et al., 2016
EmoSemio	EmoSemio	23 semantic product-specific sentences (16 positive and 7 negative emotions)	Spinelli et al., 2014
EmoSensory Profile	EMP	List of 14-17 emotion terms and 13 sensory terms.	Schouteten et al., 2015
Emotive Projection Test	EPT	Given that people tend to project their feelings onto others, emotions can indirectly be measured from judgements on personality traits of portraits of others.	Mojet et al., 2015
EsSense Profile	ESP	List of 39 emotion terms that can either be rated on 5-pt scales (EPRAT) or selected (EPCATA).	King and Meiselman, 2010
EsSense25	EsSense25	A shortened version of the EsSense Profile with 39 emotion terms.	Nestrud et al., 2016
Experimental Auction	ExpAuc	The amount of money participants bid in an auction procedure measures their willingness to pay for a certain product.	Poole et al., 2007
Free Choice Profiling	FCP	Assessors describe products in their own words and rate the perceived intensity of those terms.	Oreskovich et al., 1991
Facial Expression Response	FER	A facial expression is one or more motions or positions of the muscles beneath the skin of the face. Facial expressions can reflect the emotional state of an individual in response to a stimulus.	Bercea, 2013

Table 2-1. Continued

Method	**Abbreviation**	**Description**	**References**
Functional Magnetic Resonance Imaging	fMRI	Functional magnetic resonance imaging is a technique for measuring brain activity by detecting changes in blood oxygenation and flow that occur in response to neural activity.	O'Doherty et al., 2001
Geneva Emotion and Odor Scale	GEOS	A list with 6 scales (pleasantness, unpleasantness, sensuality, relaxation, refreshment, sensory pleasure) and 36 terms developed to investigate odor-elicited affective feelings.	Chrea et al., 2009
Hybrid Hedonic Scale	HHS	A linear scale with marked equidistant points and verbal affective labels serving as anchors in the middle and extreme regions of the scale.	Villanueva et al., 2005
Hard laddering	HL	Using an a priori defined list of paired products in combination with a structured questionnaire, assessors are asked to indicate their choice priority, and to provide arguments for their choice. [Note: In Soft Laddering the participant is interviewed by a trained experimenter.]	Russell et al., 2004
Heart Rate response	HR	Heart rate is the speed of the heartbeat measured by the number of contractions of the heart per minute (bpm). Changes in heart rate can reflect changes in the state of arousal of an individual.	Bercea, 2013
n-point Hedonic Scale (hedonic rating)	HSn	Hedonic, pleasantness or liking scales typically use (5, 7 or 9 point) category scales and uni- or bipolar magnitude estimation scales to give numerical estimates of liking.	Lim, 2011
International Positive and Negative Affect Schedule (PANAS) Short Form	I-PANAS-SF	Instrument for self-assessment of affect by rating 5 positive and 5 negative emotion terms on 5-point scales. The I-PANAS-SF is a shortened version of the PANAS.	Thompson, 2007
Implicit Positive and Negative Affect Test	IPANAT	Assessors rate to what extent artificial words (e.g., SAFME) fit with three positive (happy, cheerful, energetic) and three negative (helpless, tense, inhibited) emotions.	Quirin et al., 2009
Labelled Affective Magnitude (LAM) scale	LAM	9-pt hedonic scale with magnitude-scaled semantic labels.	Schutz and Cardello, 2001
Multi Dimensional Mood Questionnaire	MDMQ	List of 24 (long form) or 12 (short form) items covering 3 bipolar dimensions of mood (i.e., good mood-bad mood, alertness-fatigue, ease-unease).	Geier et al., 2016

Table 2-1. Continued

Method	**Abbreviation**	**Description**	**References**
Magneto-Encephalography	MEG	Magneto-encephalography is a functional neuroimaging technique for mapping brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain, using very sensitive magnetometers.	Tsourides et al., 2016
Positive and Negative Affect Schedule	PANAS	Instrument for self-assessment of affect by rating 10 positive and 10 negative emotion terms on 5-point scales.	Watson et al., 1988
Postauricular Reflex	PAR	The postauricular reflex is a vestigial muscle response in humans that acts to pull the ear backward and that increases with emotional valence.	Hebert et al., 2015
Product Choice	PC	Assessors are presented with different products and asked to select the one they prefer for consumption.	Lévy and Köster, 1999
Positron Emission Tomography	PET	A functional imaging technique that is used to observe metabolic processes in the body.	Bercea, 2013
Pick-up Latency	PL	Pick-up latency is based on the principles of approach-avoidance motivations: people are quicker to approach stimuli of positive valence than stimuli of negative valence.	Davies et al., 2012
Preference Mapping	PM	Preference mapping is a generic term involving a collection of techniques used to relate detailed sensory profiling data to consumer liking.	Clark, 1998
Profile Of Mood States	POMS	Psychological 5-point rating scale that can be used to self-assess mood states along 6 different dimensions (Tension or Anxiety, Anger or Hostility, Vigor or Activity, Fatigue or Inertia, Depression or Dejection, Confusion or Bewilderment).	McNair et al., 1971
Product Emotion Measurement Instrument	PrEmo (PrEmo2)	PrEmo is a non-verbal self-report instrument to measure different emotions (visualized by an animated cartoon character) on a 5-point scale.	Desmet et al., 2000 Laurans and Desmet, 2012
Rank Rating or Positional Relative Rating	PRR	The assessor is given all products at once and orders (ranks) them along a line in order of liking (hedonic order).	Kim and O'Mahony, 1998
Postural Sway	PS	Hedonic evaluation activates approach versus avoidance mechanisms that modulate postural sway such that pleasant stimuli elicit anterior-going sway, and unpleasant stimuli elicit posterior-going sway.	Brunyé et al., 2013
Rate-All-That-Apply	RATA	Assessors are presented with a list of sensory terms or questions and are asked to rate all those they consider applicable for describing the focal sample.	Ares et al., 2014
Repertory Grid Method	RGM	Assessors are presented 3 stimuli, first divide them in 2 similar stimuli and 1 different one, and then describe the differences between the products.	Kelly, 2003

Table 2-1. Continued

Method	**Abbreviation**	**Description**	**References**
Self Assessment Mannikin	SAM	Validated 9-point pictorial (anthropomorphic) rating scale for measuring pleasure, arousal and dominance.	Bradley and Lang, 1994
ScentMove	SM	ScentMove is a simplified version of the GEOS and consists of 6 scales (sensuality, relaxation, well-being, energy, nostalgia, disgust) with 3 terms each.	Porcherot et al., 2010
Startle Response	SR	The startle response is a largely unconscious defensive response to sudden or threatening stimuli, such as sudden noise or sharp movement, and is associated with negative affect and a state of arousal. The onset of the startle response is a brainstem reflectory reaction (startle reflex) that serves to protect vulnerable parts, such as the back of the neck (whole-body startle) and the eyes (eyeblink).	Koller and Walla, 2015
Skin Temperature response	ST	Local skin temperature changes reflect variations in blood flow in response to emotional stimuli.	Kreibig, 2010
State-Trait Anxiety Inventory	STAI	Introspective psychological inventory consisting of 40 self-report items pertaining to anxiety affect.	Spielberger, 1983
Take Away Behavior	TAB	Assessors are asked to take away any samples tested after experiments without any description.	Wichchukit and O'Mahony, 2010
Temporal Dominance of Emotions	TDE	Assessors periodically check the most dominant out of 10 emotions over the duration of the evaluation process.	Jager et al., 2014
Temporal Duration Judgement	TDJ	Time is underestimated when looking at pictures of food (compared to neutral pictures), and more so for disliked than for liked foods.	Gil et al., 2009
Visual Analog Mood Scales	VAMS	The Visual Analog Mood Scales is designed to measure 8 different general mood states (sad, happy, tense, anxious, confused, tired, energetic, irritated) on visual analog scales.	Bond and Lader, 1974
Visual Selective Attention	VSA	The allocation of visual selective attention indicated by eye fixation reflects relative preference for different products (liking).	Bercea, 2013
Word association	WA	Assessors are asked to write down the first images, associations, thoughts or feelings that come to mind.	Schmitt, 1998

Table 2-2. The toolbox table: a categorization of all 59 emotional measures extracted from our set of 101 articles.

		Emotional processing level		
		Low level (unconscious, sensory)	**Intermediate level (perceptual, early cognition)**	**High level (conscious, decision making)**
Measurement level	**Physiological**	BP, EDA, EEG (ERPs), fMRI, ST, HR, MEG, PET	EEG (frontal alpha asymmetry), fMRI, MEG, PET	N/A
	Behavioral	PAR, PS, SR	ACT, EPT, FACS, PL, TDJ, VSA	AC, TAB
	Cognitive	N/A	ExpAuc	Hedonic Scaling: HHS, HSn, LAM, PM Questionnaire with preferable foods: BUYB, BUYP, BUYR, BWS, BWSLT, FCP, HL, PC, PRR, SL*, WB Questionnaire with emotional lexicons: ASR, CATA, CD-CATA, EFT, EmoSemio, EMP, ESP ES25, GEOS, IPANASSF, IPANAT, PANAS, RATA, RGM, SM, TDE, WA Questionnaire with emotional pictures: AG, PrEmo, SAM Questionnaire with mood-related lexicons: MDMQ, POMS, STAI, VAMS

SL* (Soft Laddering) were not extracted from our inclusion criteria.

Assessment methods and their classification

To structure the measures described in the set of 101 papers we use a 3x3 framework consisting of three levels of emotional processing and three levels of measurement level as described in the next section.

As the first dimension of our 3x3 framework (processing level), we use three levels of processing: the lower level referring to unconscious and basic sensory processing, the intermediate level, referring to perception and early cognitive processing, and the higher level, referring to the conscious processing stage after cognition including decision-making and food-related behavior. This first dimension is similar to processing levels as used in SOR (Stimulus-Organism-Response) paradigm introduced by Mehrabian and Russell (1974b) and adjusted by Bitner (1992) and Lin (2004) (Schreuder et al., 2016). The second dimension of our 3x3 framework is the measurement level: physiological, behavioral, and cognitive. Physiological measures (like heart rate and skin conductivity) reflect the (largely unconscious) activity of the autonomic nervous system and bodily functions. Behavioral measures (like face and body movement and choice reaction time) reflect the unconscious and conscious responses of the body. Finally, the cognitive measures (like rating scales and questionnaires) reflect conscious opinions, choices, and decisions. This 3x3 framework provides a concise description of the different processing levels involved in the experience of multisensory environmental stimuli and their link to perceptual, emotional, and cognitive and behavioral outcomes. This framework is, therefore, well suited for our purposes.

Below, we discuss each of the nine combinations of processing and measurement levels in regard to assessing food-evoked emotions and used these nine categories to classify the 59 different measurement instruments reported.

Lower processing level (sensory processing)

When presented with food stimuli, individuals perceive and integrate information from all senses: vision, audition, taste, olfaction, and touch through unconscious, neurophysiological processes. Measures used to evaluate the emotional aspects of these processes were grouped in this category.

Physiological measures

Major peripheral physiological measures like heart rate (HR), electrodermal activity (EDA), skin temperature (ST), and blood pressure (BP) fall into this measurement level. HR is a cardiovascular measure and the most frequently used measure to evaluate emotional states as e.g. induced by viewing emotional pictures or film clips (Kreibig, 2010). However, HR has not been used much to evaluate emotions evoked by experiencing foods. One of the exceptions is a study conducted by de Wijk et al. (2012), who demonstrated that HR can indeed be used to assess food-related emotions resulting from the sight, smell, and taste of liked and disliked foods. Similar to HR, EDA has often been used to investigate how people react to viewing emotional pictures and film clips. The study by de Wijk et al. (2012) included EDA as well. Skin Temperature (ST, also referred to as finger temperature or FT) is a measure reflecting autonomic nervous system activity. Rimm-Kaufman and Kagan (1996) suggested that researchers interested in emotion might consider using ST as an informative variable recorded with infrared tele thermography. Similar to HR, Blood Pressure (BP) is also used to examine an individuals' unconscious emotional states (Kreibig, 2010). Marczinski et al. (2014) used BP and found that the consumption of energy drinks elevated BP, while it is still unclear which part of this rise was caused by the intake of nutrients and which part was caused by emotion.

In addition to these peripheral physiological measures, measures reflecting brain activity such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) can be also used to evaluate an individual's unconscious response to food stimuli at the physiological level (Bercea, 2013; Agarwal and Xavier, 2015). Event-related potentials (ERPs) are specific positive or negative peaks in the EEG following the presentation of a stimulus. An example is the P300, the size of which relates to the amount of attention given to the stimulus (Hoffman and Polich, 1998; Patel and Azzam, 2005). In addition to the P300, the late positive potential (LPP) is an ERP component that is related to stimulus control and the use of attentional resources and regulatory factors in the brain (Hajcak et al., 2009). MEG is a functional neuroimaging technique that maps electrophysiological activities inside the brain through very sensitive magnetometers (Yoshikawa et al., 2014; Tsourides et al., 2016). For instance, MEG identified a robust neural correlate of the food and non-food distinction (Tsourides et al., 2016). fMRI is another technique to evaluate more detailed activities and responses inside the brain by detecting changes in blood oxygenation and flow that occur in response to neural activity. For instance, Grabenhorst et al. (2008) showed that perceived pleasantness correlated with activity in the orbitofrontal cortex (OFC) and the pregenual cingulate cortex. More recently, Hoogeveen et al. (2015) demonstrated that older people reported higher liking ratings for sweet and salty, lower ratings for sour, and similar ratings for bitter compared to young people. Their findings indicated that these differences between younger and older adults may be associated with the reduction of right amygdala activity in older persons. PET is a functional imaging technique that is used to observe metabolic processes. Small et al. (2001) combined PET and MRI and demonstrated that different neural substrates mediate positive or appetitive and negative or aversive stimuli.

It is important to note that for every physiological sensor, several categories of physiological variables can be extracted (e.g. heart rate variability (HRV) and HR for ECG, and ERPs and power spectra for EEG). Within these categories there are usually further differences as to how the variable is defined. For instance, from subsequent RR intervals, HRV can be defined as the Root Mean Square of the Successive Differences (RMSSD), or as the power in frequency bands of interest (Veltman and Gaillard, 1998). In addition, extracting these indices relies heavily on choices with respect to time intervals across which data is examined and advanced data-processing techniques to filter, clean and classify the, often noisy, data from the physiological sensors. When using advanced analyses such as deep learning, it is not uncommon to try and compare different algorithms or parameter settings (e.g. (Saeed et al., 2017)). Finally, we want to point out the fact that physiological variables can be affected by body movements or time related factors such that failing to properly control or correct for those could lead to incorrect conclusions (Brouwer et al., 2015a). These aspects need to be kept in mind when reviewing and comparing studies using physiological measures. A detailed description and discussion of the different signal processing techniques as used in the studies using physiological measures referred to in this review is outside the scope, but literatures exist on good practice (e.g. for machine learning in the context of EEG (Lotte et al., 2007); for heart rate variability (Camm et al., 1996)).

Behavioral measures

A representative measure here is startle response (SR), also known as the alarm reaction or the startle reflex. The SR is a completely natural, involuntary reaction to a stimulus such as a flash of light, a sudden threatening movement or loud noise, and is associated with negative affect. Walla et al. (2010) provided evidence that SR modulation (eye blinks in their study) can be used reliably to quantify human motivational states related to the intake of different kinds of food (Walla et al., 2010). The postauricular reflex (PAR) is a vestigial muscle response in humans that acts to pull the ear backward and can be also grouped into this category. A study by Hebert et al. (2015) suggests that both PAR and SR are modulated by emotional states with valence scores of appetitive, neutral, and disgusting food images affecting SR in a direction opposite to PAR. In particular, pleasant stimuli enhanced the PAR (Gable and Harmon-Jones, 2009). Sandt et al. (2009) suggest that the PAR might be useful to measure appetitive responding in human. Another unconscious behavioral index categorized in this group is the postural sway (PS): a covert horizontal movement in response to a stimulus. There is strong evidence for bidirectional links between approach and avoidance (i.e. motivational state) and overt and covert indices of motor behavior (Elliot and Covington, 2001), including PS. Brunyé et al. (2013), using 100 food images, found evidence that individual preferences modulated anterior–posterior postural sway, with pleasant stimuli eliciting anterior-going sway and unpleasant stimuli elicit posterior-going sway.

Cognitive measures

The lower emotional processing level is defined as the stage in which stimuli are automatically and unconsciously processed through our senses and the brain’s sensory cortices without conscious intervention or interpretation. Cognitive measures rely on for instance individuals rating their food-evoked emotions and are thus not applicable at this level of processing.

Intermediate processing level (perception and early cognitive processing)

Following the integrated multisensory perception of food stimuli, individuals relate their percepts to previous experiences and information stored in memory. This can occur through both conscious and unconscious processes. This intermediate processing level concerns a short-term emotional state that is directly related to the object of focus. This state can be observed consciously (feeling aroused, pleasant, etc.) or can be experienced unconsciously, and drives the allocation of processing resources and priorities for the consecutive processing level (cognition, behavior, and decision).

Physiological measures

Frontal alpha asymmetry measured using EEG fits in this category. Some studies, using pictures of desserts, showed that alpha asymmetry is an unconscious response that depended on whether the subject would like to approach or avoid that dessert (Gable and Harmon-Jones, 2008; Harmon-Jones and Gable, 2009). Such a response can only occur after the perception of the dessert has been integrated with information from memory. Other neuroimaging techniques (fMRI, MEG, and PET) were also considered as measures to assess the intermediate processing level.

Behavioral measures

The autobiographical congruency test (ACT) and the emotive projection test (EPT) measure the reaction time needed to think of a happy or a sad life event and are employed to indirectly measure food-related emotions (Mojet et al., 2015). Mojet et al. (2015) showed that the ACT didn't differentiate between products, and that the EPT was the most promising measure since it had no significant correlation with either liking and differentiated between products. The temporal duration judgment (TDJ) evaluates how long someone is looking at food images. Gil et al. (2009) provided evidence that the time looked at disliked food images was longer and at liked food shorter than the time looked at neutral food images. The pick-up latency (PL) method is another indirect measure for liking based on the principles of approach-avoidance motivations: PL is smaller for positive valence and larger for negative valence (Krieglmeyer et al., 2010). Davies et al. (2012) showed that PL was reduced for positively conditioned flavors and increased for negatively conditioned flavors. Visual selective attention (VSA) is a related behavioral measure: a transitory decline in the pleasantness of the taste modulates covert visual selective attention (di Pellegrino et al., 2011). Finally, the facial expression response (FER) resulting from the integrated stimulation evoked by food experience over a short time period is a behavioral measure to assess the intermediate processing level. Several recent studies provide evidence that the FER correlates with valence and arousal ratings (de Wijk et al., 2012; Garcia-Burgos and Zamora, 2013; Danner et al., 2014b; de Wijk et al., 2014; He et al., 2016a).

Cognitive measures

This category contains instruments that do not rely on directly asking questions about the subject's emotions (as these would tap into the higher emotional processing level) but on implicit cognitions. The experimental auction (ExpAuc) technique is such an instrument, using a real product and real money (Poole et al., 2007). They showed that the ExpAuc implicitly measures an individual's willingness to pay for a certain product. A second instrument in this category is the implicit association test (IAT: (Greenwald et al., 1998)): a tool to measure implicit attitudes towards stimuli. However, no study using an IAT on food stimuli was present in our final set of articles.

Higher processing level (conscious reflection and decision making)

The higher processing level could be considered as the final stage in which individuals consciously recognize what foods are, which emotions they evoke, how these are associated with their social relationships, how food stimuli are related with their expectations, etc. The measures that evaluate these conscious emotions were categorized in this group.

Physiological measures

All of the physiological responses to food stimuli are unconscious and automatically occur in the human body and brain. Because of the implicit assumption that individuals are not able to intentionally control their physiological response to food stimuli, none of the papers selected for this review employs a measure at the physiological level to assess the higher level of emotional processing.

Behavioral measures

While the behavioral measures at the intermediate processing level reflect unconscious, short-term (immediate) emotional state, the instruments at the higher level relate to more deliberate approach or avoidance behaviors, influenced by more cognitive and long-term emotion. In this category, the measures of Amount Consumed (AC: (Zandstra et al., 1999)) and Take Away Behavior (TAB: (Weiss et al., 2010; Wichchukit and O'Mahony, 2010b)) can be included.

Cognitive measures

There are many instruments that fall within this category (more than in all eight other categories combined). Therefore, we use five subcategories to provide further structure: (1) hedonic scaling, and questionnaires with (2) preferable foods, (3) emotional lexicons, (4) emotional pictures, and (5) mood-related lexicons.

Hedonic Scaling

In Hedonic Scaling using a *n*-point scale (HSn), product evaluation (liking) is typically scored on 5-point (Brunyé et al., 2013), 7-point (Caporale et al., 2009; Awazu, 2013) or 9-point (Ares et al., 2010; Chung et al., 2012; Bhumiratana et al., 2014) liking scale. Adjectives are sometimes used to label the points in order to aid the interpretation. In most studies consumers are asked to rate several samples sequentially without reference to other elements in the set (a serial monadic approach). Some modified methods from HSn are Labelled Affective Magnitude (LAM) scale and Hybrid Hedonic Scale (HHS). Schutz and Cardello (2001) developed the LAM scale, which is 9-point hedonic scale with magnitude-scaled semantic labels. HHS is also a linear scale with marked equidistant points and verbal affective labels serving as anchors in the middle and extreme regions of the scale (Villanueva et al., 2005). Finally, Preference Mapping (PM) is a technique to describe the relationship between hedonic ratings of a randomized population of subjects and the sensory scores of the products rated by trained panels (Clark, 1998).

Questionnaires choosing preferable foods

Questionnaires asking participants to select one or more products are also based on their higher processing level of food stimuli. Instruments in this subcategory are the Product Choice (PC) test, where assessors are presented with different products and asked to select the one they prefer for consumption (Lévy and Köster, 1999), the Free Choice Profiling (FCP) test, where assessors describe products in their own words and rate the perceived intensity of those terms (Kim et al., 2013), and the Positional Relative Rating (PRR) test. Kim and O'Mahony (1998) used the latter measure, in which assessors are given all products at once and order them along a line in order of liking. The Hard Laddering (HL) method also belongs in this category. In a laddering task, assessors are asked to compare products or their attributes and elicit their reasons for choosing a certain product for purchase or consumption. In HL, a structured questionnaire is applied, while in soft laddering the participant is interviewed by a trained experimenter (Russell et al., 2004). There was no study using the soft laddering technique selected by our inclusion criteria. In addition to using HL and studying their Buying Behavior (BUYBeh), Rosas-Nexticapa et al. (2005) also asked participants to rate their Buying Preference (BUYPref) and Buying Ranking (BUYRank). They demonstrated that these ratings might predict purchase frequency of products over a one year period of experiments. However, this type of study is too costly and time-consuming to be practical, and these rating scales do not reflect a person’s actual buying behavior in all circumstances as shown by Lange et al. (2002). The last measures in this subcategory are the Best-Worst Scaling (BWS: (Jaeger et al., 2008) and the Best-Worst Scaling of Lexicon Terms (BWSLT: (Thomson et al., 2010)).

Questionnaires with emotional lexicons

King and Meiselman (2010) compiled a list of 39 emotional terms that consumers associate with products, known as the EsSense Profile (ESP). Each of these terms is rated on a 5-point scale. When applied to evaluate food products, the ESP provides additional information that is not explained by overall product liking (King et al., 2010). A shortened version of ESP, named EsSense 25 (compiling 25 emotional terms) was developed later (Nestrud et al., 2013). Also, Spinelli et al. (2014) developed a product-specific questionnaire, called EmoSemio, based on one-on-one interviews conducted with a modified version of the Repertory Grid Method (RGM: (Kelly, 2003)). They provided evidence that EmoSemio discriminated product specific emotions better than ESP with chocolate and hazelnut spreads as samples (Spinelli et al., 2014). The Scent Move (SM: (Porcherot et al., 2010)) is also a tool using an emotional lexicon (a simplified version of the Geneva Emotion and Odor Scale or GEOS: (Porcherot et al., 2010)) as is the Check-All-That-Apply (CATA) technique, in which assessors are presented with a list of sensory emotional terms or phrases and are asked to select all those terms or phrases they consider applicable to describe the focal sample (Adams et al., 2007). The modified CATA measure to evaluate food-evoked emotions is the Consumer-Defined Check-All-That-Apply (CD-CATA), and was developed and demonstrated by Ng et al. (2013). The Rate-All-That-Apply (RATA) is a rating-based variant of CATA (Ares et al., 2014). In the Word Association (WA) technique, assessors are asked about concepts, images and thoughts that come into their mind for each product, yielding thoughts and associations about the products, after conscious evaluation. The Temporal Dominance of Emotions (TDE) tool is based on the Temporal Dominance of Sensations (TDS), which evaluates the sequence of dominant sensations of a product during a certain time period (Pineau et al., 2003), but with emotional instead of sensory attributes (Ares et al., 2008). They showed that temporal emotional attitude was related to the sensory profiles obtained with TDS. Other self-reported emotion questionnaires were also included in this category, such as EmoSensory Profile (EMP) rating (a combination of 14-17 emotional terms and 13 sensory terms: (Schouteten et al., 2015)), Affect Self Report (ASR) scale (rating 18 affective terms on a 7-point scale: (Christie and Friedman, 2004)), Empathic Food Test (EFT; rating 12 empathic terms: (Geier et al., 2016)), Positive and Negative Affect Schedule (PANAS; rating 10 positive and 10 negative emotion terms: (Watson et al., 1988)), Implicit Positive, Negative Affect Test (IPANAT; rating 3-positive and 3-negative emotions: (Quirin et al., 2009)), and International Positive and Negative Affect Schedule Short Form (I-PANAS-SF: (Thompson, 2007)), a shortened version of the PANAS (Watson et al., 1988; Thompson, 2007)).

Questionnaires with emotional pictures

The Self-Assessment Mannikin (SAM) developed by Bradley and Lang (1994) is a measure to evaluate emotions (valence, arousal, and dominance) that uses pictures instead of text (as do the emotional lexicons described above). Similar to the SAM, the Affect Grid (AG: (Russell et al., 1989)) and PrEmo (PrEmo-2; (Desmet et al., 2000b; Laurans and Desmet, 2012b)) were also developed for participants to more easily describe their emotions with pictures. Visually expressed emotions are hypothesized to more closely resemble intuitively experienced emotions (Dalenberg et al., 2014). Evidence for this hypothesis stems from EEG-experiments showing that emotion processing is faster for facial expressions than for emotional words (Schacht and Sommer, 2009; Frühholz et al., 2011; Rellecke et al., 2011).

Questionnaires with mood-related lexicons

The final subcategory consists of self-report techniques that evaluate mood, such as the Multi Dimensional Mood Questionnaire (MDMQ) that employs 24 (long form) or 12 (short form) items to cover three bipolar dimensions of mood (Geier et al., 2016), the Profile Of Mood States (POMS) that uses ratings of 6 mood states along with 6 different dimensions (McNair et al., 1971), the State-Trait Anxiety Inventory (STAI) that uses ratings of 40 self-report items pertaining to anxiety affect (Spielberger, 1983), and the Visual Analog Mood Scales (VAMS) that employs ratings of 8 different general mood states (Bond and Lader, 1974).

Discussion

Emotions are considered to be important drivers of food-related cognitions and behavior like food choice and eating behavior. Indeed, Dalenberg et al. (2014) provided the results indicating that the predicting power of individual's food choice got better by adding the evaluation of food-evoked emotions with liking rating scores. As Köster and Mojet (2015b) state, there is no doubt that unconscious emotions can play a role in eating and drinking behavior in a way that is independent of hedonic pleasure as measured by liking. In addition, also mentioned in the introduction, some studies provided the results that liking scaling do not predict individual's actual food choice (Zandstra and El-Deredy, 2011; Griffioen-Roose et al., 2013a). Valid, reliable and sensitive instruments that assess food-evoked emotions are therefore valuable for fundamental and applied research and for instance in developing new food products and advocating a healthy lifestyle. A complicating factor in this field is that human emotion is a multifaceted construct linked to physiological, behavioral, and cognitive processes, and we may not assume to find a single measure that covers the full range, while there is a general conviction that all facets are relevant.

We listed, organized, and reviewed the prevailing instruments based on a literature review consisting of 101 peer-reviewed articles published between 1997 and 2016. Our main observations are: (1) There is an overabundance of different measures (about 59 in our set of 101 papers); (2) The majority of these measures assess the cognitive level of emotional processing using subjective ratings or questionnaires (i.e., self-reports are over-represented); (3) Articles that report two or more measures generally use measures that all tap into the same level of emotional processing, while it may be expected that redundant measures have limited added value.

The overabundance of measures

As also mentioned in the introduction, the fact that about 59 different measures are employed in 101 papers makes it evident that there is not a 'golden standard' to assess food-evoked emotions. This has consequences regarding the validity of the measures, the generalizability of reported effects, and the integration (or even meta-analysis) over studies. From a methodological point of view, the use of uncommon measures of which the validity is unknown is not desirable. Some measures are developed and exclusively used by few research groups. For instance, previous studies developed questionnaires with emotional lexicons for a range of product categories (chocolate in (Thomson et al., 2010); blackcurrant squashes in (Ng et al., 2013); chocolate and hazelnut spreads in (Spinelli et al., 2014); coffee in (Bhumiratana et al., 2014)). Those questionnaires with lexicons are specific for each product and cannot be applied for universal food products. Few other research groups use this technique with the same emotional lexicons. We recommend to choose widely applied, validated measures whenever possible. One should not construct one's own instrument before having verified that there is not an existing tool that may serve one's goals. Both our Tables 2-1 and 2-2 may be of assistance here.

The over-representation of self-report measures

Of the 59 measures reported, eight are at the physiological, 11 at the behavioral, and 40 at the cognitive measurement level. More than 60% of the reported measures are based on self-reports assessing cognitive emotional processing. Among them, more than 80% studies used a HSn measure as one of the measures. There are two important comments to be made here. First, as also discussed in the introduction, self-reports (although successful) have inherent shortcomings: emotions are difficult to verbalize, the ‘emotional’ lexicon varies across cultures and languages, answers may be biased, and verbalizing emotions can interfere with food experience itself. Second, self-reports assess almost per definition the higher levels of emotional processing and cannot assess unconscious emotional processing, while this is deemed important to improve both our understanding of food-evoked emotions as well as our algorithms to predict future cognition and behavior. As indicated from the toolbox table with 3x3 framework provided here (Table 2-2), most of the papers selected by our inclusion criteria are of a mono-disciplinary nature and contribute more to the perpetuation of the narrow tunnel view of the cognitive measurement level.

The limited added value of redundant measures

Although authors often report the data of two or more measures, these are almost always of the same category, again reflecting the over-representation of self-report measures: the majority of those studies used a combination of HSn (including liking and pleasantness scale) ratings and questionnaires with emotional lexicons. These redundant tests with comparable results can be useful to prove the robustness of a specific effect such as hedonic asymmetry. Hedonic asymmetry refers to the finding that people overwhelmingly use positive rather than negative words, whether describing recalled food experiences or describing reactions to food samples. Hedonic asymmetry was described by Desmet and Schifferstein (2008) and has been replicated for different types of commercial products by using ESP and LAM (Cardello et al., 2012), ESP and HSn (King and Meiselman, 2010), CD-CATA and HSn (Ng et al., 2013), and with the original emotional lexicons and HSn (Desmet and Schifferstein, 2008; Manzocco et al., 2013). High correlation between tests was also found by Cordonnier and Delwiche (2008) who reported that PRR yielded similar results to HSn with lemonades as stimuli. Similarly it has been reported that LAM has equal reliability and sensitivity to HSn and a somewhat greater discrimination ability among highly liked foods than HSn (Schutz and Cardello, 2001). These results indicate that, although redundant instruments showing the same results may improve the robustness of the findings, the added value of their repeated use may be limited and the scope of the conclusions is necessarily restricted to that of the chosen measurement category. This limited added value may not outweigh the extra costs of the test and the burden subjected to the participants. In case the redundant instruments show contradictory effects, this could be a useful indication for the lack of robustness and replicability of the effect, but one should carefully consider the quality of the data and the validity of one or both tests. We recommend to refrain from using instruments from the same category. We do recommend using multiple tests but to choose tests from different categories (see below under toolbox for guidance on how to choose your combination of tests, that should preferably be along the diagonal of Table 2-2). Although these different measures sometimes provided complementary information, there were also many cases in which they yielded redundant and sometimes even contradictory results.

A toolbox table to support the selection of a combination of instruments

To structure the 59 measures, we used nine categories based on a 3x3 framework with the dimensions: measurement level (physiological, behavioral, and cognitive) and emotional processing level (low - unconscious sensory processing, intermediate - perception and early cognitive processing, and high - conscious decision-making and behavior). Furthermore, the category 'cognitive high processing level' was divided into five different subcategories. The resulting classification (see Table 2-2) indicates that current physiological methods are used to evaluate the low and intermediate processing levels, behavioral methods to evaluate all three processing levels, and cognitive methods to evaluate the intermediate and high processing levels. The resulting classification or 'toolbox table' (see Table 2-2) can be used to select a minimal set of methods that provides maximal (complementary) information on the aspects of affective food experience that are of interest, for instance by choosing methods along the table's diagonal from top-left to bottom-right. Current widespread practice is to use one or more measures from the bottom-right category only. Although such measures can certainly provide much information about the consciously experienced effects of food, they may miss essential nuances in feelings and emotions that may influence later behavior. Non-verbal, implicit measurements may complement verbal self-report questionnaires to better understand individual's food-evoked emotions. Although the number of tools in some of the framework's cells is still limited, new instruments have recently been developed and tested. Examples include facial movements, such as smacks of mouth and lips and tongue protrusion for hedonic reactions and gape, eye quench, and nose wrinkle for aversive reactions (Steiner et al., 2001), facial expressions (Kostyra et al., 2016), ANS responses such as HR and EDA as signals for negative emotions like disgust and anger that also provided detailed information on food preference (de Wijk et al., 2012). Using machinelearning techniques, Brouwer et al. (2017a) recently found that a combination of different physiological measures showed whether a participant was cooking and tasting a dish with conventional ingredients (a chicken stirfry) or a dish with high arousal, low valence ingredients (a mealworm stirfry).

In addition to these successful demonstrations, some studies conducted a correlation analysis between self-reports questionnaires and physiological or behavioral measurements (Garcia-Burgos and Zamora, 2013); (Mojet et al., 2015). Garcia-Burgos and Zamora (2013) conducted linear regression analysis between self-reported hedonic value and the intensity of disgust facial expression analysis on bitter-tasting foods and found a quite low correlation value. The fact that the different categories indeed seem to measure different facets of emotional processing implies that combining them may increase our understanding (for example of the discrepancies between hedonic ratings and consumer choice) and ultimately improve our predictions of food-related cognitions and behavior taking into account that food-evoked emotions is only part of a range of factors that influence future liking, choosing or buying behavior. It was also proposed that the implicit nature of food-related behavior requires the development of more appropriate/adequate research methods that measure the motives of the consumer and her reactions to food in a more implicit way. We recommend acquiring more accumulative and simultaneously collected physiological, behavioral, and self-report datasets.

Implications for related research

Viewing images of food triggers the desire for the real thing: just looking at pictures of food causes salivation (Spence, 2011b) and an uptick in ghrelin, a hormone that causes hunger (Schüssler et al., 2012). These effects increase when images represent food in a more vivid way (Spence, 2011b; Moore and Konrath, 2015). Vividness (Steuer, 1992), also referred to as media richness (Daft and Lengel, 1986) refers to the sensory breadth (the number of sensory dimensions) and sensory depth (the information quality and resolution) of stimuli. Vivid stimuli allow observers to fill in more missing sensory information and thereby diminish the user' s perception of mediation (i.e., the indirect perception of a product through technical means or devices). This enables users to activate a fuller, more concrete or vivid mental model of a mediated product, which in turn affects their product appraisal (Choi and Taylor, 2014) and intensifies the imagined product experience (Roggeveen et al., 2015). It has for instance been found that vivid (full color) images of pizza elicited higher levels of food craving, a stronger salivation response, and stronger eating intentions, than similar pallid (black and white) images (Moore and Konrath, 2015). Also, vivid food cinemagraphs evoke stronger appetitive responses than similar stills (Toet et al., 2019d). Virtual reality (VR:(Gorini et al., 2010a; Nordbo et al., 2015a; Ung et al., 2018b)) and augmented reality (AR: (Narumi et al., 2012a; Pallavicini et al., 2016)) appear to be promising tools to study the impact of environmental cues on human nutritional behavior since they typically provide vivid imagery. In the field of Human-Computer Interaction (HCI), novel multisensory (taste, smell, tactile) interfaces are being developed and used to support studies on food-related emotions and behavior, personal health and wellbeing (Comber et al., 2014a; Obrist et al., 2016), or simply to enhance or augment the experience of food (Narumi et al., 2011; Schöning et al., 2012; Spence and Piqueras-Fiszman, 2014; Velasco et al., 2018a). HCI can promote healthy food practices and social dining experiences (Comber et al., 2013). The addition of tactile and olfactory channels to VR and AR systems will further enhance the vividness of the mediated perception, and may for instance afford shared distributed virtual multisensory dining experiences (Braun et al., 2016). Multisensory technologies also allow researchers to control the various inputs that accompany a given food experience (Velasco et al., 2018a). In all the aforementioned applications it is essential to monitor the emotional and behavioral responses to the perceived virtual or mediated food, and to assess how these responses compare to those evoked by real food (Pallavicini et al., 2016). Reliable emotion assessment techniques are therefore required to further develop and optimize multisensory interactive experiences and to assess the ecological validity of mediated food presentations (Obrist et al., 2016; Obrist et al., 2017).

Conclusions and recommendations

Common practice in assessing food-evoked emotions relies on many different instruments of which the majority only assess a limited aspect of emotional processing. While this common practice resulted in robust and relevant findings, the plethora of different instruments hampers the validation of infrequently used instruments and the comparison of results over studies. The restriction to assess only the cognitive emotional processing level makes it difficult to obtain a complete picture and impoverishes our predictive models. We recommend the following:

(1) use widely applied, validated measures and refrain from constructing a new tool unless absolutely necessary,

(2) refrain from using (highly correlated) instruments from the same category,

(3) use multiple measures from different categories, preferably covering all three emotional processing levels, for instance by selecting tests along the diagonal of Table 2-2.

(4) acquire and share simultaneously collected physiological, behavioral, and cognitive datasets to improve the predictive power of food choice and other models.

It must be noted that there are many more factors to fill in the gap between individual's food-evoked emotions and food choice and buying behavior in the supermarket, such as food culture, habitats, incomes, and family structure in addition to food-evoked emotions. However, finding even more accurate measures or proper combination of measures to better interpret individual's food-evoked emotions is a definite step to better predict individual's actual food choice and buying behavior.

Acknowledgements

The authors thank Dymphie van der Heyden for her assistance with the search for relevant papers in the PsycINFO database.

Chapter 3

Explicit and Implicit Responses to Tasting Drinks Associated with Different Tasting Experiences

This chapter is published as:
Kaneko, D., Hogervorst, M., Toet, A., Van Erp, J.B.F., Kallen, V., and Brouwer, A.M. (2019). Explicit and implicit responses to tasting drinks associated with different tasting experiences. Sensors 19, 4397.

Abstract

Probing food experience or liking through verbal ratings has its shortcomings. We compare explicit ratings to a range of (neuro)physiological and behavioral measures with respect to their performance in distinguishing drinks associated with different emotional experience. Seventy participants tasted and rated the valence and arousal of eight regular drinks and a “ground truth” high-arousal, low-valence vinegar solution. The discriminative power for distinguishing between the vinegar solution and the regular drinks was highest for sip size, followed by valence ratings, arousal ratings, heart rate, skin conductance level, facial expression of “disgust,” pupil diameter, and Electroencephalogram (EEG) frontal alpha asymmetry. Within the regular drinks, a positive correlation was found between rated arousal and heart rate, and a negative correlation between rated arousal and Heart Rate Variability (HRV). Most physiological measures showed consistent temporal patterns over time following the announcement of the drink and taking a sip. This was consistent over all nine drinks, but the peaks were substantially higher for the vinegar solution than for the regular drinks, likely caused by emotion. Our results indicate that implicit variables have the potential to differentiate between drinks associated with different emotional experiences. In addition, this study gives us insight into the physiological temporal response patterns associated with taking a sip.

Introduction

Information about food-evoked emotions in addition to simple liking ratings have been argued to improve predictions regarding consumers' food choices (Kuenzel et al., 2010; Thomson et al., 2010; Cardello et al., 2012; Dalenberg et al., 2014; Gutjar et al., 2015c). Researchers have developed and used emotion-association questionnaires, in which individuals indicate to what extent they experience certain emotions after tasting foods and/or beverages (Köster and Mojet, 2015b; Schifferstein, 2015; Kenney and Adhikari, 2016a). Such explicit self-reporting measures are relatively easy and cost-effective to apply. However, they have inherent drawbacks: they are discontinuous, prone to demand characteristics, may suffer from response biases, and may not cover subconscious processes (Venkatraman et al., 2015; Kaneko et al., 2018a). Furthermore, the "emotional" lexicon varies across cultures and languages, and consumers are not used to verbalizing their emotions, particularly when it comes to foods (Gutjar et al., 2015c). Finally, when consumers are asked to explicitly rate their emotions, this can interfere with the food experience itself (Winkielman et al., 2011a). Several authors propose to measure unconscious (implicit) responses in addition to self-reports in order to better understand consumers' food-evoked emotions and predict their future food choice behavior (Köster and Mojet, 2015b; Beyts et al., 2017; Verastegui-Tena et al., 2017). These measures can be of a (neuro)physiological nature (e.g., brain signals) or behavioral (e.g., facial expression). The current study aims to provide an overview of the sensitivity of a range of simultaneously measured implicit and explicit measures in response to tasting drinks that are associated with different affective experiences.

Several non-verbal, implicit measurements have been studied before in the context of probing affective food experience in response to tasting. With respect to facial movements, Steiner et al. investigated affective reactions to pleasant and unpleasant food tastes in human infants and primates (Steiner et al., 2001). In their experiment, they clearly distinguished between hedonic and aversive expressions without the use of questionnaires. More recently, Danner et al. conducted facial expression analysis elicited by six different fruit and vegetable juices and obtained a negative association of the facial expression ''neutral'' with disliking and a positive association of ''angry'' and ''disgusted'' expressions with disliking, indicating that facial expression analysis may complement self-report questionnaires (Danner et al., 2014b). Similar results were found in (de Wijk et al., 2014) which reported more neutral facial expressions were elicited by liked breakfast drinks compared to less liked ones.

A few studies examined physiological responses to tasting in the context of affective food experiences. In these studies, participants tasted basic taste solutions, juices, and foods while recording the autonomic nervous system through skin conductance or electrodermal activity (EDA) and heart rate (HR) (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b). The drinks and foods were also explicitly rated with respect to their liking or pleasantness. They all reported that high heart rate and high EDA was associated with unpleasant ratings.

In emotions research, emotions are often described not only in terms of pleasantness or valence but also in terms of arousal, or the intensity of the emotion (or (neural) activation) (Russell and Pratt, 1980; Lang, 1995; Berridge, 1996; Berridge et al., 2010; Vansteelandt et al., 2013; Piqueras-Fiszman et al., 2014). While valence and arousal are in principle independent, the drink and food stimuli, as used in References (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b), may have confounded pleasantness (valence) and arousal, where unpleasant food or drinks were high in arousal and vice versa. This would fit with the notion that especially EDA is a reliable indicator of arousal (the sweat glands being innervated by the sympathetic part of the autonomic nervous system) rather than valence (Roth, 1983). In (de Wijk et al., 2014), autonomic nervous system responses elicited by tasting different breakfast drinks were recorded, and participants were not only questioned about liking, but also about intensity. As could be expected, positive correlation coefficients between intensity and EDA were found. There was a negative correlation between intensity and heart rate. In contrast to a high EDA and high heart rate for unpleasant ratings as reported in the studies mentioned above (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b), positive correlation coefficients between liking and EDA, as well as between liking and heart rate, were found in (de Wijk et al., 2014). However, from this paper, it is not clear whether all these correlations were significant. The differences in ratings between drinks were very small.

As indicated by the literature reviewed above, relating implicit physiological and behavioral measurements to emotional food experience is not straightforward. In general, relations between neurophysiology and emotion depend on both stimuli and context (Cacioppo and Gardner, 1999; Kreibig, 2010; Brouwer et al., 2015c). It is still an open question as to what extent implicit (neuro)physiological measures, facial expressions, and behavioral measures can be used to monitor emotional food experience relative to, or in addition to, explicit self-report measures, and how they compare to one another. Existing studies used only a few implicit measures, and most asked only for liking or preference scores, or types of emotion, thus omitting the arousal dimension.

The current study aims to fill these gaps in the literature by simultaneously examining a wide range of implicit measures and by including arousal ratings. Furthermore, we examine the case in which individuals know what they are about to taste. This is usually the case in daily life, but contrasts with most studies that focus on the effect of taste only and therefore do not provide any other information about the drink or food (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b; Gomes et al., 2016). Finally, while most studies used stimuli either associated with very strong or very subtle differences in emotional experience, we take both approaches in this study. This is done because both approaches have their drawbacks as well as their merits. Since explicit ratings can be biased, it is difficult to assess the “real” emotional experience. This is referred to as the ground truth problem (Brouwer et al., 2015c). From this point of view, it is a good choice to use a stimulus that is a priori known to be associated with a strong emotion. This will result in choosing a quite extreme stimulus, such as a quinine solution (Steiner et al., 2001), or in our case, a vinegar solution. On the other hand, we do not know whether findings from studies with an extreme stimulus generalize to more subtle differences. In the current study, we therefore explore whether examining responses to regular drinks with subtle differences in emotional experience on the one hand, and comparisons between regular drinks and a strongly emotional “ground truth” drink on the other hand, lead to similar results with respect to the sensitivity of the investigated variables to reflect emotional food experience. Below, we shortly outline our study and define specific hypotheses.

In the present study, participants were informed what drink they were about to taste before they took a sip of the drink and rated the drink's valence and arousal. Besides these explicit measures of valence and arousal, we recorded a range of implicit measures while the participants were performing this task: sip size, facial expression of disgust, neurophysiological measures (Electroencephalogram (EEG)), and measures of the autonomic nervous system (pupil size, EDA, and HR). As mentioned above, we take two approaches to probe the sensitivity of these measures of emotional food experience. The first approach is to include a special drink that is expected to be more strongly associated with a certain emotion than the other regular drinks. We chose a drink that evokes a high-arousal, low-valence affective response (a vinegar solution). Comparing the responses to the regular drinks to the response to this generally disliked ("ground truth") drink indicates the discriminative power of a specific response measure, i.e., how well the different response measures distinguish a strongly disliked drink from regular drinks (with associated emotions that are close to each other). Second, we tested the sensitivity of implicit measures to distinguish responses to the regular drinks that only differed slightly in terms of associated emotions. We use self-reported valence and arousal as generally accepted measures of the affective experience of regular drinks, and correlate each of the remaining measures with these traditional measures.

We expect the high-arousal, low-valence experience of the vinegar solution to be reflected in the explicit ratings and a smaller sip size. For the rest of our measures, we hypothesize the following associations. Because arousal is consistently and positively related to both pupil size and EDA (Roth, 1983; Bradley et al., 1992), we expect a larger pupil size and increased EDA when tasting the vinegar solution compared with the regular drinks. While no straightforward relation between heart rate and arousal or valence exists across contexts (Brouwer et al., 2015c), the previous studies suggest a higher heart rate for high arousal in the context of tasting (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b). We also expect heart rate variability (HRV) to be negatively associated with stress or arousal, as reported in other contexts (Berntson et al., 1997; Grossman and Taylor, 2007). For valence, we expect the facial expression of disgust to be informative (Steiner et al., 2001; Danner et al., 2014b; de Wijk et al., 2014). Furthermore, we examine frontal brain activation as an indicator of valence. Relatively strong frontal left brain activation has been associated with positive valence and relatively strong right activation with negative valence (Ohgami et al., 2006; Brouwer et al., 2017c). Research reviewed by Harmon-Jones et al. (Harmon-Jones et al., 2010) indicates that rather than valence, frontal brain asymmetry more parsimoniously maps onto approach and avoidance, where relatively strong right frontal brain activation, as indicated by frontal EEG alpha asymmetry, is associated with avoidance and the reverse with approach motivation. In general, negative valence can be associated with both approach (e.g., being angry and wanting to fight) and avoidance motivation (e.g., being scared and wanting to flee). In the food domain, however, high valence can be expected to be associated with approach motivation and low valence with avoidance motivation. We therefore view alpha asymmetry as an indicator of valence. As far as we know, this measure has not been studied in tasting, but there is evidence for its relation with the approach/avoidance motivation or valence coming from studies using food pictures (Gable and Harmon-Jones, 2008; Harmon-Jones and Gable, 2009) and cooking and tasting a chicken versus mealworm dish (Brouwer et al., 2017c).

In sum, the research questions in this study are (i) how well do different self-reported, physiological, and behavioral variables discriminate regular drinks from a drink that is known to be strongly disliked, (ii) how sensitive are these measures to the subtle differences between the regular drinks, and (iii) how are different implicit measures associated with self-reported valence and arousal for the regular drinks? In addition to these research questions, we examine the general temporal pattern of physiological variables after the announcement of the drink and taking a sip.

Materials and Methods

Participants

A total of 70 healthy participants (19 men, 51 women) took part in this study. All of them were of Dutch nationality and were between 19 and 63 years old, with an average of 48.5 years and a standard deviation of 10.5 years. Participants were recruited through the participant pool of the research institute where the study took place (TNO Netherlands Organisation for Applied Scientific Research (TNO)) and received a monetary reward to compensate for time and travel costs. All participants signed an informed consent in accordance with the Helsinki Declaration of 1975, as revised in 2014 (WorldMedicalAssociation, 2014), before participating in this study. The study was approved by the TNO Institutional Review Board. Three participants were excluded due to technical problems related to the registration of event markers and physiological data recording. This left us with 67 participants for further analysis. For the analysis of facial expression, we only investigated participants without glasses (42).

Materials

Recording Equipment

EDA (for skin conductance level—SCL), ECG (for inter-beat interval—IBI), and EEG (for frontal alpha asymmetry—FAA) were recorded using an Active Two MkII system (Biosemi B.V., Amsterdam, The Netherlands), with a sampling frequency of 512 Hz. SCL was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. ECG electrodes were placed on the right clavicle and on the lowest floating left rib. For EEG, 32 active silver-chloride electrodes were placed according to the 10-20 system.

Pupil diameter (PD) was recorded at 60 frames per second using SmartEyePro V6.1.6 (Smart Eye AB, Göteburg, Sweden). This system consists of two cameras (Basler acA640-120gm, HR 8.0 mm lens) placed at the left and right side of a screen that presented the name of the drink and the rating scales. The screen had a size of 37.0 by 30.0 cm and the viewing distance was approximately 80.0 cm.

Participants' faces were recorded using a Color CCTV Camera, WV-CP150E (Panasonic Corp., Osaka, Japan) during the entire experiment. The video camera was positioned at the left side of the screen.

Self-Report Rating Scales and Sip Size

The self-assessment manikin (SAM) (Bradley and Lang, 1994) with nine-point scales were used for valence and arousal self-report ratings. The nine-point scale was positioned in the appropriate location at the bottom of each SAM, where the most leftward (most unpleasant and calm) and the most rightward (most pleasant and aroused) parts of the scale corresponded to values of 1 and 9, respectively. With respect to valence, participants were asked how pleasant their experience with the drink was, with the manikin on the right indicating a very pleasant experience and the manikin on the left a very unpleasant experience. With respect to arousal, participants were asked how intensely they experienced the drink, with the manikin on the right indicating a very intense experience and the manikin on the left a very calm experience. Also, they were instructed that they should try to answer quickly, without thinking too long.

For the behavioral measure of sip size, the exact weight of each drink including the cup was measured before the participant took a sip. After finishing the experiment, the cups containing the rest of the drinks were weighted again to determine the sip size.

Samples

The drinks used in this study were apple juice (Appelsientje), orange juice (Appelsientje), yogurt drink (Vifit), milk (Campina), buttermilk (Campina), rooibos tea (Pickwick), black tea (Pickwick), cola (Coca-cola), and diluted vinegar (Private Brand of Plus: 50% vinegar, 50% water) solution. The regular drinks were chosen to represent a variation in basic flavors and temperature. They differed from one another in taste, but they were expected to be close to one another in affective experience, at least relative to diluted vinegar. Teas and the vinegar solution were always prepared in the same way each morning. Teas were kept at about 60 °C, and the vinegar solution was kept at room temperature. The other drinks were kept in a refrigerator before being served to the participants. Each sample was served in a white plain cup, in portions of 50 g. Participants tasted the drinks in randomized order except for a 50 g cup of water, which was always presented after the vinegar solution to decrease the possible lingering of emotional and physiological effects. Responses to water are not included in the analyses.

Design and Procedure

After participants arrived at the laboratory, the experimental procedure was explained, and they were asked to sign the informed consent form. The electrodes for EDA, ECG, and EEG were attached, and participants were asked to sit comfortably in front of the screen. The EDA electrodes were worn on the non-dominant hand, and participants were asked to pick up the sample cups with their dominant hand. Participants filled out a general questionnaire on demographic details and current emotional state. Before the experiment started, the SmartEyePro system was calibrated.

Then, the experimenter showed and explained how to take a sip, immediately putting the cup down after the sip, and participants performed a practice trial with water. After this there was time for additional practice or instructions when needed, and participants had the chance to ask questions. The timeline of an experimental trial is indicated in Figure 3-1. First, the name of the drink was presented on the screen. This was the sign for the experimenter to place the appropriate drink in front of the participant. After 5 seconds, the name of the drink disappeared, which was the sign for participants to take one sip.

After taking the sip, participants sat still and looked at a blank white screen. Forty seconds after the name of the drink had appeared on the screen, the self-report valence and arousal rating scales appeared, with the valence scale on top. After rating valence and arousal, the name of the next drink appeared on the screen. This procedure was repeated until all drinks had been served.

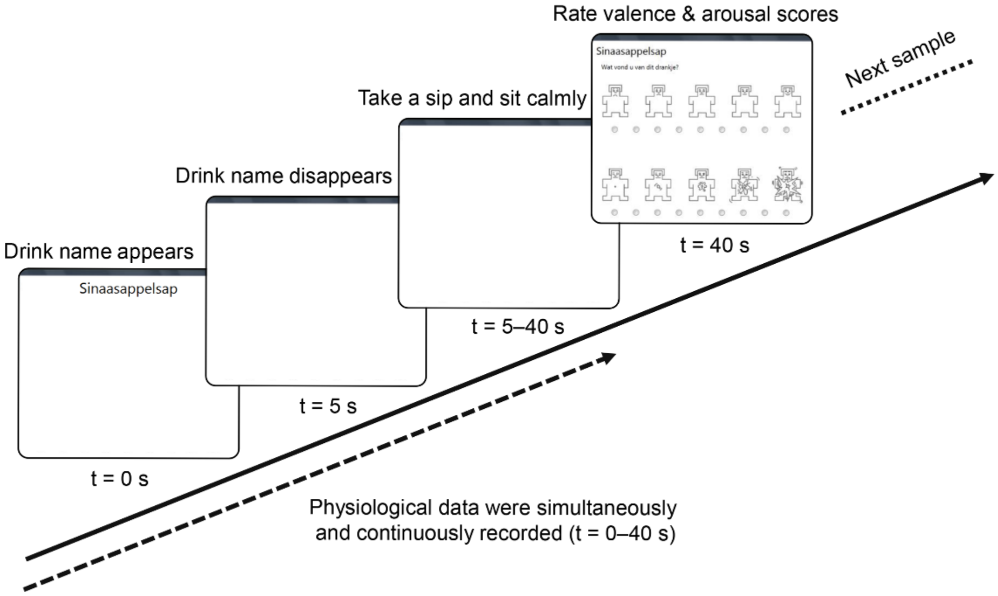


Figure 3-1. Schematic overview of an experimental trial. Participants clicked the small circles below the SAM scales in order to give their response, after which, a new trial started.

Data Processing and Analysis

The analysis performed on the physiological data for EDA, ECG, and EEG were similar to analyses in the previous studies (Hogervorst et al., 2014; Brouwer et al., 2015b; Brouwer et al., 2017c).

Preprocessing for Facial Expression, EDA, ECG, and EEG

For the analysis of the facial expression of disgust ($FR_{disgust}$), the video data was analyzed using FaceReader software version 7.0 (Noldus Information Technology B.V., Wageningen, The Netherlands) at a sampling rate of 12.5 frames per second. FaceReader extracts the basic emotional expressions, including disgust, using an artificial neural network that was trained on over 10,000 pictures and exploiting a number of facial features, including facial action units, gaze direction, and head orientation. Calibration procedures were conducted for each participant to correct for person-specific biases toward a certain facial expression according to the FaceReader manual. $FR_{disgust}$ is expressed as a value from 0 to 1 in each frame, indicating the intensity of the emotion. “0” means that the emotion is not visible in the facial expression, “1” means that the emotion is fully present.

The EDA signal was bandpass filtered between 0.03 and 100 Hz. Inter-beat interval (IBI), defined as the temporal distance between R-spikes (Appelhans and Luecken, 2006), was extracted from the ECG signal using custom made MATLAB 2019a (www.mathworks.com) algorithms.

Raw EEG data were pre-processed and analyzed using MATLAB and the FieldTrip open source MATLAB toolbox (Oostenveld et al., 2011). The EEG pre-processing entailed standard procedures of referencing the signals to the average EEG signal and filtering them using a 0.5 Hz high pass and a 43 Hz low pass filter to remove slow drifts and high-frequency noise. Logistic infomax independent component analysis (ICA, (Bell and Sejnowski, 1995)) was performed to classify artifactual independent components, i.e., components not reflecting sources of neural activity, but were rather ocular or muscle-related artifacts. These components were removed from the data. This was done using EEGLAB v14.1.2 for MATLAB (Delorme and Makeig, 2004). Measurement intervals, starting at the onset of the announcement of the drink and ending 40 seconds later (at the time that the rating scales appeared), were divided into 5 s intervals. For each of these intervals, the spectral power was calculated over bands ranging from 8 to 13 Hz (alpha) in steps of 0.2 Hz following a fast Fourier transform (FFT) approach using a single Hanning taper. Subsequently, values were integrated. FAA at F7 and F8 was determined for each 5-s segment by taking the relative difference between alpha as recorded at the right and the left side of the cortex: $((R - L) / (R + L)) \times 100$ (Papousek et al., 2014). Positive values indicate lower alpha power in the left than in the right hemisphere (i.e., relatively greater left hemisphere cortical activity).

Extraction of Variables

For each of the variables (valence, arousal, sip size, $FR_{disgust}$, SCL, IBI, HRV, PD, and FAA), we required one value for each participant and each drink. For rated valence, rated arousal, and sip size, one value was already present. For the continuously measured variables, these values were extracted as follows: For $FR_{disgust}$ and IBI, we averaged values across the forty seconds starting at onset of the announcement of the drink and ending 40 seconds later. Since PD and FAA were rather noisy, we used the median across the 40 seconds rather than the mean. HRV was calculated as the root of the mean of squared successive differences of the IBIs (RMSSD) across the 40 seconds (i.e., the average absolute difference between successive IBIs (Camm et al., 1996; Goedhart et al., 2007)). SCL showed strong drifts across the duration of the experiment. Therefore, before taking the average across the 40 second intervals, SCL curves were baselined using the average of the first 5 seconds of data. Subsequently, sip size, SCL, IBI, and PD were log transformed, leading to more normal distributions.

To remove irrelevant overall differences between participants, we centered the data by subtracting the mean value for each variable and each participant.

For each variable, before and after this subtraction, values that were more than five standard deviations away from the mean were discarded as outliers. For FAA, this was also done for the alpha values that were used to compute the FAA. This procedure led to 9.9% lost data for HRV, 5.3% lost data for sip size, 10.4% lost data for FAA, and less than 2.0% of data loss for all other variables.

To examine how the different physiological variables evolved over time, we also determined a value for each participant and drink in the same way as described above, but for each subsequent 5-s interval rather than the whole 40s. To visualize potential differences in patterns between drinks clearly, we baselined the curves for each drink using the first 5 s of the data. This was not done for HRV, since 5 s was too short an interval to obtain HRV in a meaningful way.

Statistical Analysis

To examine how well different variable discriminated regular drinks from the vinegar solution, we calculated one z-score (z) for each variable using the equation below:

$$z = \frac{(\text{mean value regular drinks}) - (\text{mean value vinegar solution})}{\sigma_{diff}}$$

To calculate sigma (σ_{diff}), we started with the distribution of the values in response to the vinegar solution, and the distribution of the values in response to all other eight drinks. The standard deviation of the distribution of the differences between these two was estimated by taking the width of the 95% confidence interval of the difference and dividing this by 4. The z-score (or discriminative power) was significant (at $p < 0.05$) when larger than 1.96.

Next, we performed individual one-way ANOVAs for each variable with the regular drink as the independent variable (eight regular drinks). These tests indicated how sensitive the measures were to the subtle differences between the regular drinks. While a large F-value and a low *p*-value were indicative of a sensitive measure, it would make a difference whether this hinged upon only one or several significant comparisons. Therefore, for the measures that showed a significant effect for a "regular drink," we also report how many and which of the total possible number of 28 pairwise post-hoc comparisons between regular drinks reached significance (Tukey's HSD).

To investigate the association between different implicit measures and self-reported valence and arousal for the regular drinks, we calculated (for each implicit measure separately) the correlation between the implicit measures and both valence and arousal. We used the scores averaged per drink as input, i.e., one implicit value and one rating score per drink, resulting in eight data pairs. This analysis explored whether there was a systematic ordering of the drinks along both dependent variables.

Results

Sensitivity of Measures to Distinguish between Regular Drinks and Diluted Vinegar

Reported valence and arousal of each drink, averaged across participants, are shown in Figure 3-2. The valence and arousal ratings of the vinegar solution were on average the lowest and the highest of all drinks tested, respectively. These results are in accordance with our assumption that the vinegar solution could serve as the ground truth unpleasant and arousing drink.

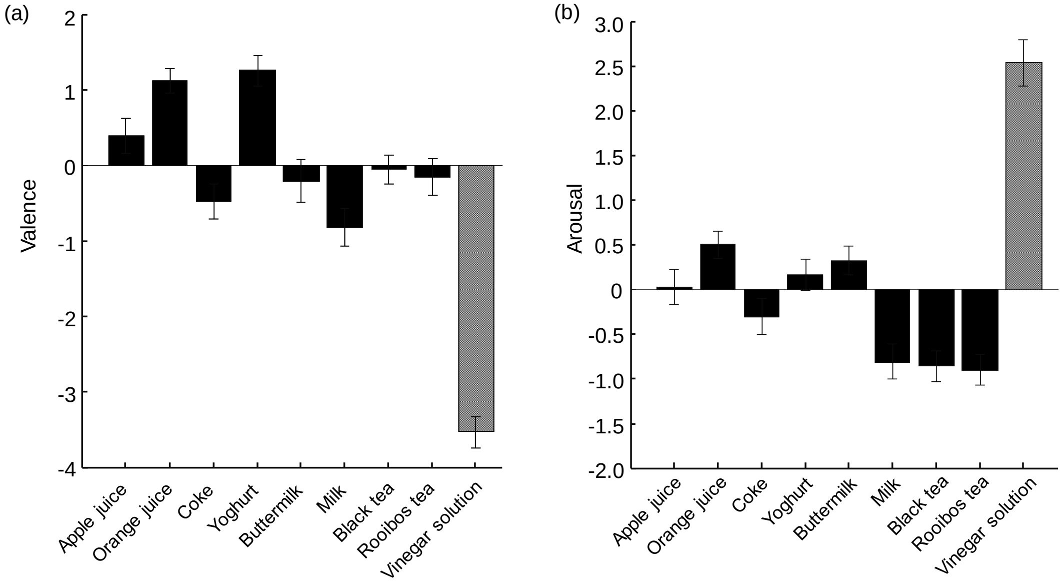


Figure 3-2. The results of explicit ratings: valence (**a**) and arousal (**b**) averaged across participants. Solid bars represent the eight regular drinks, and the dotted bar represents the vinegar solution.

The implicit behavioral measures, sip size, and $FR_{disgust}$ are shown in Figure 3-3, separately for each drink and averaged across participants. The sip size of the vinegar solution was the smallest, and the $FR_{disgust}$ was the highest.

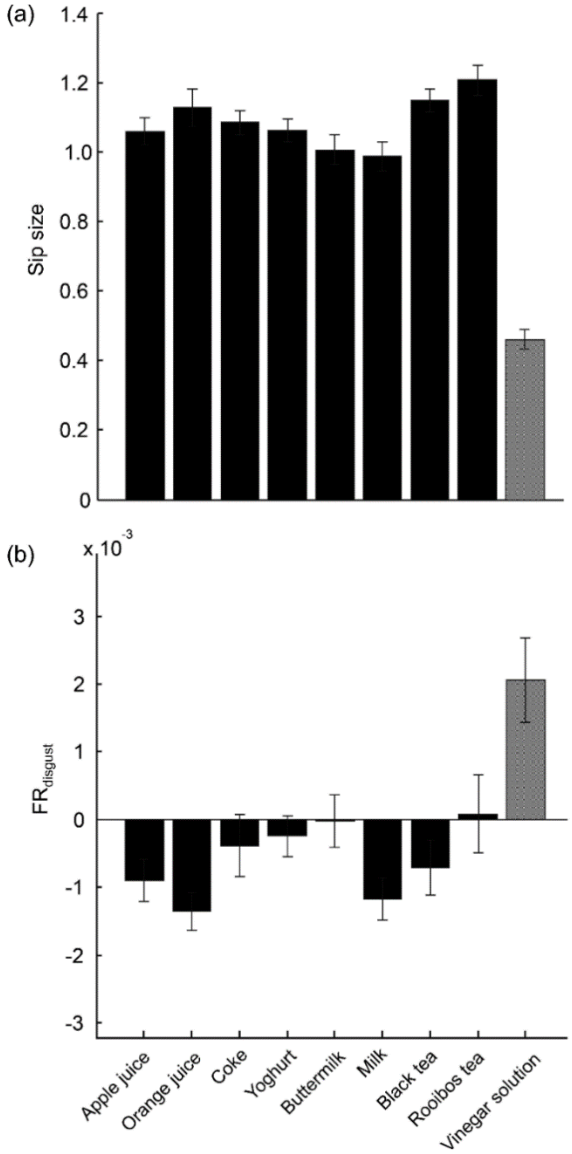


Figure 3-3. The results of behavioral measures: sip size (**a**) and $FR_{disgust}$ (**b**) averaged across participants. Solid bars represent the eight regular drinks, and the dotted bar represents the vinegar solution.

Figure 3-4 shows SCL, IBI, HRV, PD, and FAA for each drink averaged over participants. All measures show the most extreme value for the vinegar solution in the expected direction: it was the highest of all drinks for SCL and PD, and the lowest for IBI, HRV, and FAA (low FAA indicating a higher alpha power in the left rather than in the right hemisphere, i.e., relatively greater right hemisphere cortical activity, consistent with negative valence or avoidance).

The discriminative power (z-score) to distinguish vinegar from regular drinks is presented in Table 3-1. Sip size had the highest discriminative power of all measures, followed by the explicit ratings of valence and arousal, and IBI, all with z-scores higher than 10. $FR_{disgust}$, SCL, and PD had z-scores between 4 and 8, indicating highly significant discriminative power. For HRV, the z-score was below 1.96, indicating that the discriminative power of this measure to distinguish between the vinegar solution and the regular drinks was too low to reach significance.

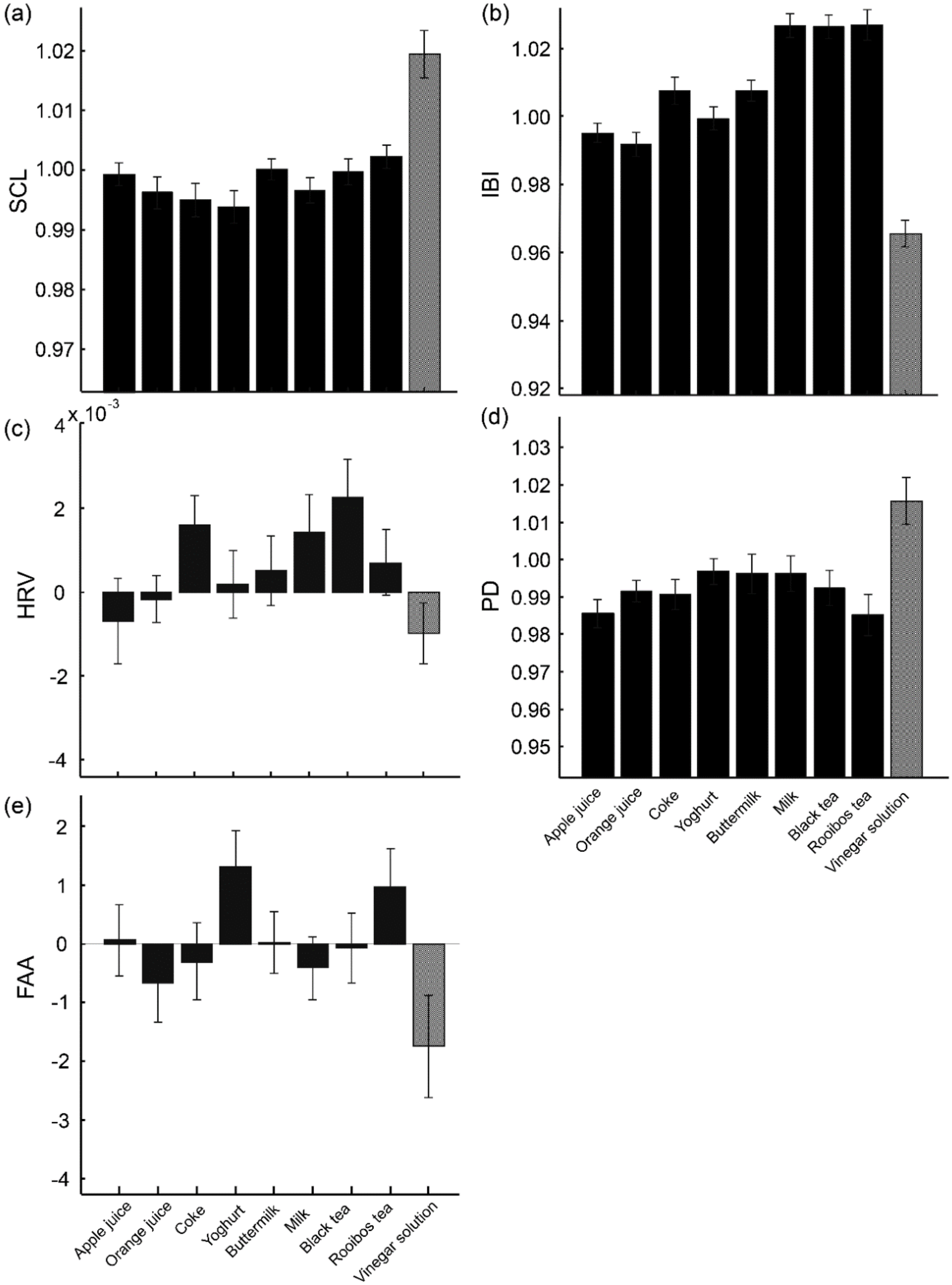


Figure 3-4. The results of physiological measures: SCL (**a**), IBI (**b**), HRV (**c**), PD (**d**), and FAA (**e**) averaged across participants. Solid bars represent the eight regular drinks, and the dotted bar represents the vinegar solution.

The discriminative power (z-score) to distinguish vinegar from regular drinks is presented in Table 3-1. Sip size had the highest discriminative power of all measures, followed by the explicit ratings of valence and arousal, and IBI, all with z-scores higher than 10. $FR_{disgust}$, SCL, and PD had z-scores between 4 and 8, indicating highly significant discriminative power. For HRV, the z-score was below 1.96, indicating that the discriminative power of this measure to distinguish between the vinegar solution and the regular drinks was too low to reach significance.

Table 3-1. Z-scores of each measure between vinegar solution and the regular drinks.

Measures	**Discrimination Power: Vinegar vs Regular Drinks (z)**
Valence	14.84
Arousal	13.42
Sip size	19.50
$FR_{disgust}$	5.97
SCL	7.97
IBI	10.99
HRV	1.94
PD	4.33
FAA	2.63

Sensitivity of Measures to Distinguish between Regular Drinks

For each separate measure, a one-way repeated measures ANOVA with regular drink as independent variable was conducted to evaluate its sensitivity to distinguish between regular drinks. In these analyses, data associated with tasting the vinegar solution were left out. The results are summarized in Table 3-2. Valence and arousal ratings, sip size, and IBI showed significant responses regarding regular drinks. $FR_{disgust}$, SCL, HRV, PD, and FAA did not differ significantly between drinks. The sensitivity of the measures as indicated by the F- and p-values were paralleled by the number of significant comparisons, as indicated by the post hoc tests. Among all 28 possible pairwise combinations of the 8 regular drinks, 11, 13, 2, and 17 combinations were significantly different to each other in terms of valence, arousal, sip size, and IBI, respectively. Table 3-3 presents which pairs of drinks differed for which measure.

Table 3-2. F and *p*-values of each measure with ANOVAs reflecting the effect of the eight regular drinks.

Measures	**Regular Drink Sensitivity (F)**	**Regular Drink Sensitivity (*p*)**
Valence	10.6	<.001
Arousal	10.1	<.001
Sip size	3.10	.003
$FR_{disgust}$	1.83	.080
SCL	1.51	.162
IBI	17.0	<.001
HRV	1.31	.244
PD	0.87	.532
FAA	1.16	.327

Table 3-3. Combinations of regular drinks that were significantly different ($p < 0.05$) according the post-hoc tests in terms of valence (“V”), arousal (“A”), sip size (“S”), and IBI (“I”). Examples of no significant differences for any of the four measures is indicated with “ns.”

	Apple Juice	**Orange Juice**	**Milk**	**Buttermilk**	**Yogurt Drink**	**Coke**	**Rooibos Tea**	**Black tea**
Apple Juice	-	ns	V, A, I	ns	ns	ns	A, I	A, I
Orange Juice		-	V, A, I	V, I	ns	V, A, I	V, A, I	V, A, I
Milk			-	A, I	V, A, I	I	S	ns
Buttermilk				-	V	ns	A, S, I	A, I
Yogurt Drink					-	V	V, A, I	V, A, I
Coke						-	I	I
Rooibos Tea							-	ns
Black Tea								-

Association between Implicit Measures and Self-Reported Valence and Arousal for Regular Drinks

Table 3-4 summarizes the results of the correlation analyses. Cells where we hypothesized a correlation based on the literature (see Introduction) are highlighted in grey. Valence was significantly correlated with IBI. Arousal was significantly correlated with IBI and HRV. Figure 3-5 presents the data underlying the three significant correlations.

Table 3-4. The summarized correlation analysis between explicit ratings (valence and arousal) and behavioral (sip size and $FR_{disgust}$) and physiological measures (SCL, IBI, HRV, PD, and FAA). The cells highlighted in light gray represent the correlations we hypothesized based on the physiological literature. The bold data represent significant correlations ($p < 0.05$).

	Valence		Arousal	
	ρ	*p*-value	*ρ*	*p*-value
Sip size	0.2290	0.585	−0.3051	0.462
$FR_{disgust}$	−0.2285	0.586	−0.0720	0.866
SCL	−0.3488	0.397	−0.3718	0.365
IBI	**−0.7326**	**0.039**	**−0.9225**	**0.001**
HRV	−0.6413	0.087	**−0.7093**	**0.049**
PD	0.0189	0.965	0.2210	0.599
FAA	0.3570	0.385	0.6180	0.102

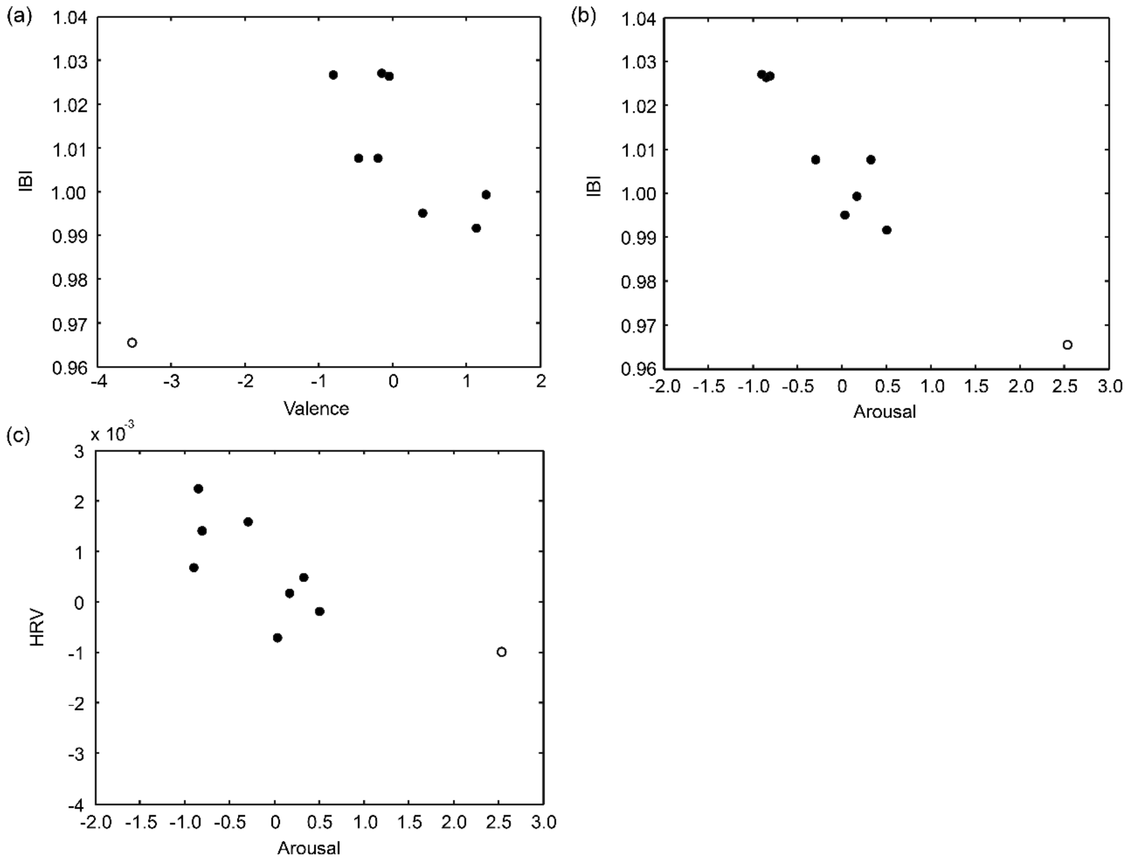


Figure 3-5. The data plots displaying the significant correlations between (**a**) valence and IBI, (**b**) arousal and IBI, and (**c**) arousal and HRV. Solid circles represent the eight regular drinks, and the open circle represents the vinegar solution. Note that correlation analysis were performed on the eight regular drinks only.

General Temporal Pattern of Behavioral and Physiological Variables Associated with Taking a Sip

The continuous measures ($FR_{disgust}$, SCL, IBI, PD, and FAA) are plotted over time in Figure 3-6. The onset of the announcement of the drinks occurred at $t = 0$ s, and the rating scales appeared at $t = 40$ s. Consistent patterns arose for almost all variables. Every time a participant started taking a sip ($t = 5$ s), $FR_{disgust}$, SCL, and PD increased, and IBI decreased for all drinks. These patterns could be partly due to effects of movements and ingestion processes associated with picking up a cup, taking a sip, and putting it down. However, consistent with the results presented in the preceding sections, these increases and decreases were clearly stronger for the vinegar solution compared to the regular drinks. For FAA, we did not see a consistent pattern over time. The difference between the vinegar solution and the regular drinks in FAA arose immediately after presentation of the name of the drink. Thus, after aligning the curves at $t = 0$, the vinegar solution does not stand out as it does for the other continuous measures presented in Figure 3-6.

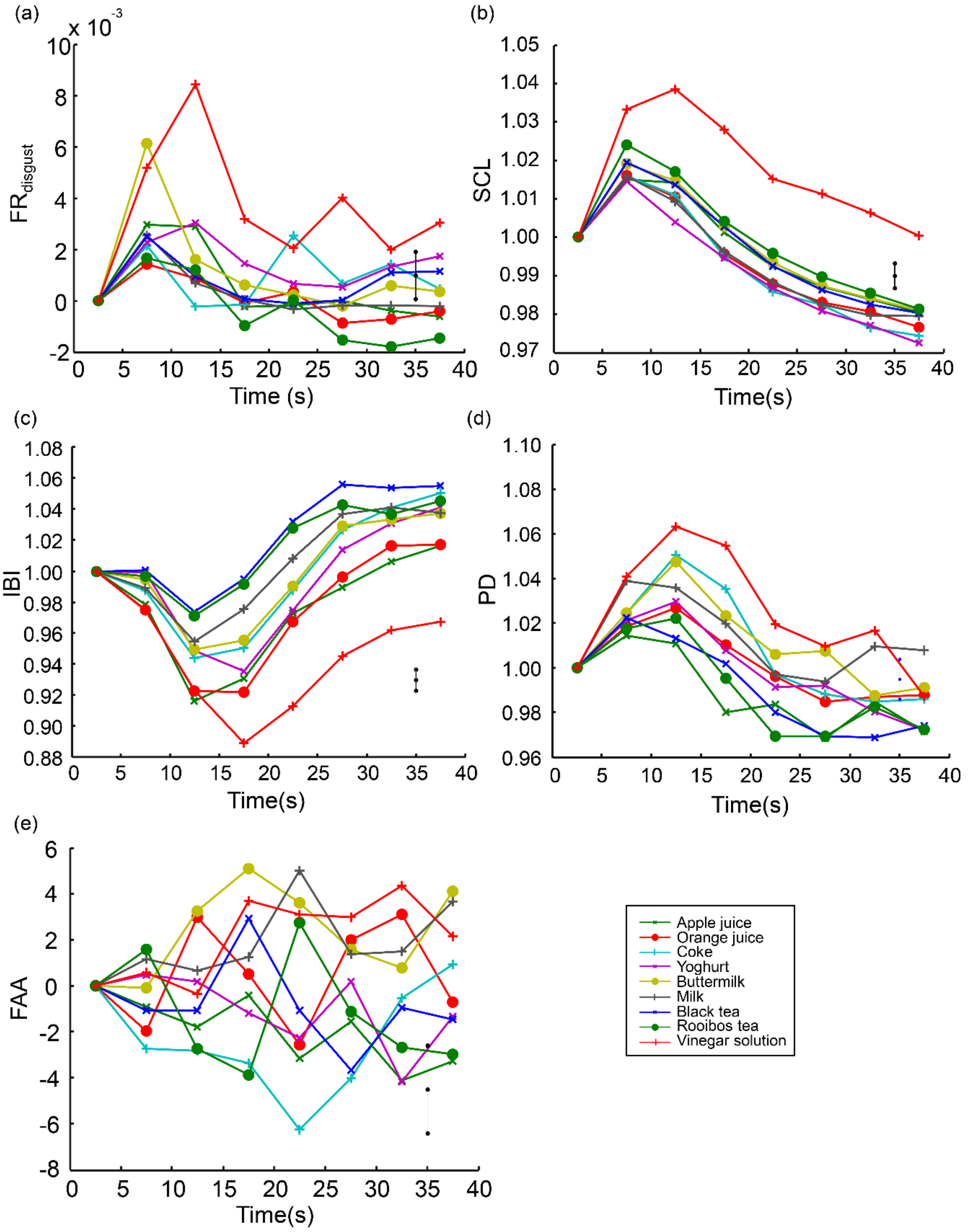


Figure 3-6. The continuous behavioral measure $FR_{disgust}$ (**a**) and the physiological measures SCL (**b**), IBI (**c**), PD (**d**), and FAA (**e**) plotted over time from the time that the name of the drink appeared (t = 0 s) to just before the rating scales appeared (t = 40 s), averaged across participants and separately for each drink.Discussion

Discussion

The present study evaluated nine different measures of emotional food experience: explicit measures (valence and arousal ratings), implicit behavioral measures (sip size and facial expression of disgust), implicit physiological measures (SCL, IBI, HRV, and PD), and an implicit neurophysiological measure (FAA). We recorded these measures while participants took sips of eight different regular drinks, and one non-regular drink (diluted vinegar). The vinegar solution was expected to differ strongly from the regular drinks in the associated emotional experience by producing the lowest valence and the highest arousal. Our data indeed showed that participants rated the vinegar solution lowest in valence and highest in arousal, took the smallest sip, and showed the most outspoken signs of disgust in their facial expression. We also found, as expected, that the vinegar solution led to higher SCL and PD, lower IBI and HRV, and FAA in the direction of avoidance (negative valence) when compared to the regular drinks.

The first research question of this study concerned the extent to which explicit and implicit measures could discriminate the vinegar solution (as a ground truth high-arousal, low-valence drink) from the regular drinks. We used z-scores as an index of discriminative power. Sip size had the highest discriminative power, even higher than the explicit valence and arousal ratings. IBI and SCL also showed high discriminative power. Although the scores for $FR_{disgust}$ and PD are somewhat lower, they are still highly significant. FAA seemed to be a less strongly discriminative measure than $FR_{disgust}$ and PD, but it still reliably distinguished between the regular drinks and the vinegar solution. Only HRV did not significantly discriminate between the vinegar solution and the regular drinks. Thus, in addition to explicit ratings, a range of implicit measures could be useful parameters to measure individual's emotions evoked by a food experience, at least for cases in which food experiences differ strongly.

For regular drinks, the effect of a drink in the ANOVAs on explicit ratings of valence and arousal suggested that participants also agreed on small differences in affective experience. This enabled us to answer our second research question about the sensitivity of the different measures to reflect subtle differences in affective experience. ANOVAs on sip size and IBI showed that, like explicit ratings, these are sensitive measures as well. $FR_{disgust}$ and SCL did not reach significance, and HRV, PD, and FAA were also not sensitive enough to detect the minor differences between the regular drinks. These results are in line with the results on discriminative power to separate the vinegar solution from the other drinks as discussed above.

The third research question concerned the association between implicit measures and self-reported explicit ratings of the regular drinks. The correlation analyses on the average scores for each drink revealed significant and high correlations ($\rho > 0.70$, explained variance > 50%) between IBI and both valence and arousal, and between HRV and arousal. The correlations between IBI and arousal, and between HRV and arousal, were the correlations that we expected to find based on the literature, and they were in the expected direction. Remarkably, HRV showed significant correlations with explicit ratings while it did not show the effects of each drink tested through the ANOVAs, in which explicit ratings were not taken into account. The fact that HRV did not distinguish well between the vinegar solution and the regular drinks, while this measure did correlate with the explicit rating of arousal for the regular drinks, can be understood when observing the position of vinegar solution in the scatter plot (Figure 3-5c): the relation between HRV and the explicit measure of arousal did not extend from the regular drinks to vinegar solution. Thus, in contrast to IBI where the relation between explicit arousal and IBI extended from the regular drinks to the vinegar solution (Figure 3-5b), HRV did not seem to be a valid marker of the affective experience of drinks that are associated with extreme levels of affective experience.

Fourth, we examined the response pattern over time of the continuous measures to provide insight into the specific patterns before, during, and after tasting. This is important in the context of extracting dependent physiological and facial expression variables and designing research on tasting that includes physiological measures. For this, we needed to know what interval length was suitable to examine physiological data relative to the time of a sip, and how much time should preferably be allowed between sips.

We found that being presented with the name of a drink and taking a sip produces characteristic patterns in most continuous variables for all nine drinks. This may be due to movement and ingestion related processes, and affective components that may always occur when taking a sip in the experiment. The distinction between the regular drinks and the ground truth high-arousal, low-valence vinegar solution was not reflected in the pattern itself but in the fact that the pattern was more distinct for the vinegar solution compared to the regular drinks: a stronger increase in $FR_{disgust}$, SCL, and PD, and a stronger decrease in IBI. We found that it took 10 to 15 seconds for these differences to fully develop. Few previous studies examined the pattern of physiological variables over time following a sip. Rousmans et al. show a few example traces of EDA and HR following the intake of a taste solution (Rousmans et al., 2000), and de Wijk et al. show patterns averaged across participant per breakfast drink over 8 seconds following the instruction to take a sip (de Wijk et al., 2014). In both studies, HR increased and then decreased again. For EDA, de Wijk et al. show a decrease across 8 seconds, while Rousmans et al. reports an increase.

The general increase in HR (i.e., the decrease in IBI) after taking a sip that we found here was smaller than that found in (de Wijk et al., 2014) and (Rousmans et al., 2000). De Wijk et al. show increases of about 12 bpm for all drinks (de Wijk et al., 2014). In (Rousmans et al., 2000), increases vary between 1.3 (water) and 11 (quinine sulfate). In (de Wijk et al., 2014), movements were minimized more than in (Rousmans et al., 2000) and the current study. The participants in (de Wijk et al., 2014) sat still with a straw in their mouth and on a signal, took a sip and kept sitting still until it was indicated that they could take the straw out of their mouth. In (Rousmans et al., 2000), participants took a sip from a cup themselves, similar as our participants. The fact that the strongest increases were found in an experiment where participants sat relatively still suggests that the general increase of HR does not seem to be mainly caused by the movement of the hand (e.g., holding a cup and taking a sip).

We evaluated nine explicit and implicit potential measures of affective food experience. For all of them, we found at least some evidence of their sensitivity. The nine measures differ in several ways and the preferred (combination of) measure(s) will depend on the research question at hand. Explicit valence and arousal ratings have a good ability to measure both large and subtle differences in emotions, but have several disadvantages, as described in the introduction. Sip size had the largest power to discriminate the ground truth of a low-valence and high-arousal vinegar solution from the regular drinks. Although it is not a continuous measure, it taps into implicit behavior and may thus be less prone to response biases than the explicit ratings of valence and arousal. IBI has the advantage of being both continuous and implicit, and did consistently well in all tests; it appeared to be sensitive to both large and subtle differences in affective experience. The correlation between valence and IBI for the regular drinks indicated a high HR for high valence, adding to the mixed findings on associations between HR and valence in tasting studies, as discussed in the introduction. We argue that rather than a relation between HR and valence, the relation in this context was actually between HR and arousal. The correlation between arousal and IBI is much stronger than between valence and IBI, and valence and arousal were not independent in the stimuli we used. We substantiated this idea by performing additional partial correlation analyses, showing that the correlation between IBI and valence disappeared when controlling for arousal, while the correlation between IBI and arousal remained when controlling for valence. $FR_{disgust}$, SCL, and FAA may not be sensitive enough to easily identify subtle differences but were definitely able to discriminate large differences. Apart from being continuous, they were not correlated to the ratings of valence and arousal for regular drinks, and may be considered as tapping into a fundamentally different dimension than these explicit measures. Finally, HRV turned out to be not suitable to discriminate between vinegar solution and regular drinks. However, when analyzing the regular drinks without considering the vinegar solution, HRV did show a significant correlation with explicit ratings of respectively arousal and valence, and may help to increase the validity and reliability of rank ordering regular drinks with subtle emotional differences along arousal and valence scales.

There are some limitations in this study. For each of the physiological and facial expression data streams, choices were made as to which variable to extract and in what way. We aimed to represent each data stream by one (a priori) promising variable, but it may be that other extraction methods or variables show different (better) results. Also, adding a long resting baseline after answering the questionnaire and before the appearance of the name of the drink may have resulted in less noise and higher sensitivity of the physiological variables. In our experiment, participants were asked to take a sip themselves in an effort to enhance the naturalness of tasting. However, the downside of this is added noise through movement and short, partial occlusion of the face when the cup is at the mouth. We here examined the situation that participants tasted a drink that they expected to taste (as is common in daily life). The food experience and physiological processes that we examined were therefore the result of a mixture of expectation and sensory processes, starting at the moment that the name of the drink appears on the screen. We refer to Verastegui-Tena and colleagues for physiological studies that specifically look at the role of expectation in tasting (Verastegui-Tena et al., 2017). A final limitation we want to mention is the fact that rated valence and arousal of the regular drinks correlated positively; including regular drinks that are high in arousal and low in valence, or low in arousal and high in valence, would help to disentangle valence and arousal effects.

Conclusions

In this study, we tested regular drinks varying in sweetness, carbonation, temperature, sourness, and thickness that were expected to differ slightly with respect to associated affective experience, as well as one “ground truth” low-valence, high-arousal drink, to evaluate the potential of different explicit and implicit measures to reflect food experience. This resulted in a comprehensive overview of the sensitivity of each of the measures to reflect different affective food experiences strongly, as well as more subtle differences. Furthermore, we showed the association between explicit measures and different implicit measures. Out of the complete set of implicit measures (sip size, facial expression of disgust, skin conductance level, heart rate, heart rate variability, and EEG frontal asymmetry), heart rate showed good sensitivity in all cases. We argue that heart rate should be viewed as a measure of arousal rather than valence. Finally, we provided insight into the development of continuous implicit variables over time after taking a sip of drinks differing in affective experience. Our results may guide the design of future studies and applications utilizing implicit measures for quantifying affective experience, which may ultimately enable the continuous monitoring of food experience without influencing the experience itself.

Acknowledgements

The authors thank Esther Bosch, Wouter Korteling, Ivo Stuldreher, and Benjamin de Graaff for their assistance with conducting the experiment and analyzing data.

Chapter 4

Estimating affective taste experience using combined implicit behavioral and neurophysiological measures

This chapter is published as:
Brouwer, A.M., Van Den Broek, T., Hogervorst, M., Kaneko, D., Toet, A., Kallen, V., and Van Erp, J. (2020). Estimating affective taste experience using combined implicit behavioral and neurophysiological measures. IEEE Transactions on Affective Computing.

Abstract

We trained a model to distinguish an extreme high arousal, unpleasant drink from regular drinks based on a range of implicit behavioral and physiological responses to naturalistic tasting. The trained model predicted arousal ratings of regular drinks, highlighting the possibility to estimate affective experience without having to rely on subjective ratings.

Introduction

Knowledge of individuals' affective experience associated with eating or drinking certain foods or drinks is of great interest to the food industry as well as other parties involved in developing food products (Prescott, 2017) and those promoting healthy eating behaviour. As reviewed in (Lagast et al., 2017; Kaneko et al., 2018a), affective food experience is usually measured using verbal questionnaires. While this is a relatively cheap method with high face validity, it suffers from certain drawbacks. Verbal descriptions are discontinuous, prone to demand characteristics, may suffer from response biases, and may not cover subconscious processes (Venkatraman et al., 2015; Lagast et al., 2017; Kaneko et al., 2018a). Also, the 'emotional lexicon' strongly varies between individuals (Gutjar et al., 2015b; Prescott, 2017) and slight variations in phrasing can have (unintended) effects on the results (Toet et al., 2018a). Finally, repeatedly asking individuals about their food experience will interfere with the food experience itself, different from what is experienced in a natural, daily life situation. Measuring unconscious (implicit) responses in addition to self-reports may help us to better understand consumers' food-evoked emotions and possibly predict their future food choice behavior (Köster and Mojet, 2015b; Verastegui-Tena et al., 2017).

Physiological, implicit measures have been studied before in relation to food experience elicited by tasting (Lagast et al., 2017; Kaneko et al., 2018a). High heart rate and skin conductance or electrodermal activity (EDA) have been reported for unpleasant tastes (Rousmans et al., 2000; de Wijk et al., 2012; Danner et al., 2014b) though (trends of) the opposite have been found as well (de Wijk et al., 2014). From a physiological point of view, variables reflecting activity of the autonomic nervous system such as EDA, heart rate and also pupil diameter and heart rate variability are expected to be more closely associated with arousal rather than pleasantness (valence) (Mendes, 2009). The reported relations between valence and these variables may have been mediated by arousal, where valence and arousal may have been associated in the stimuli used. While existing tasting studies often omit the arousal axis of affect, de Wijk et al. (2014) included ratings of arousal. They indeed found the expected positive relation between EDA and arousal ratings.

A variable that could be expected to convey valence is EEG (electroencephalography) frontal alpha asymmetry. This measure has been shown to be related to approach-avoidance motivation (Harmon-Jones et al., 2010), which in the context of food, may be considered as a proxy for valence. Expected alpha asymmetry effects have been found in viewing and cooking food (Gable and Harmon-Jones, 2008; Brouwer et al., 2017a) but, as far as we are aware, not in tasting.

We recently conducted a study with the aim of comparing these measures all in the same context of tasting drinks from cups (Kaneko et al., 2019a). We designed our experiment to mimic natural tasting as well as possible. This means that rather than employing strict laboratory conditions and rigid sampling regimes (e.g. (Rousmans et al., 2000; de Wijk et al., 2012; de Wijk et al., 2014)), participants tasted common drinks, they knew what drink they were about to taste, and apart from being instructed to take one sip at a designated time point, and to take sips in a similar way, they were not instructed in other ways (e.g. on exactly how long to keep the sip in the mouth). With this approach, we cannot disentangle effects of different sensory processes (vision, olfaction and taste) and expectation, and it will introduce more noise due to (slightly varying) body movements. On the other hand, it is more compatible with a real-life tasting situation. Also, this naturalistic approach gained us the potentially interesting, implicit variable of sip size. In the remainder, with 'tasting', we refer to the complete set of sensory and anticipatory effects that go with tasting a known drink. We refer to Verastegui-Tena and colleagues for physiological studies that specifically look at the role of expectation in tasting (Verastegui-Tena et al., 2017). In our study, participants were asked to taste eight regular drinks as well as a vinegar solution, that was a priori expected to be strongly disliked (high arousal and low valence compared to the regular drinks). We recorded the physiological implicit measures as discussed above, the implicit behavioural measure of sip size, and ratings of valence and arousal. These were all tested for their sensitivity to distinguish between the regular drinks and the vinegar solution. Physiological measures were also examined with respect to their development over time. EDA, heart rate and pupil diameter showed consistent temporal patterns: increasing after presentation of the name of the drink which was the sign to take one sip, peaking between 5-15 sec and then decreasing again. The discriminative power of distinguishing between the vinegar solution and the regular drinks as expressed by a z-score was highest for sip size (z=19.5), followed by rated valence (z=14.8), rated arousal (z=13.4), heart rate (z=11.0), EDA (z=8.0), pupil diameter (z=4.3) and EEG frontal asymmetry (z=2.6). There was no significant difference between vinegar solution and regular drinks for heart rate variability. For the regular drinks, positive correlations were found between arousal ratings and heart rate, and between rated arousal and heart rate variability (where the heart rate variability recorded during tasting the vinegar solution was at approximately the same level as the drink next highest in rated arousal, suggesting that a floor level may have been reached that prevented this variable from successfully distinguishing between the vinegar solution and the regular drinks). For regular drinks, there was also a positive relation between valence and heart rate, but further data inspection and analyses indicated that this was probably mediated by a positive relation between rated arousal and valence in regular drinks.

These results were encouraging with respect to the potential of implicit measures to estimate affective food experience. In the current study, we used the data of this same study to take a more quantitative approach and explore the potential of combining implicit measurements, covering the behavioral, physiological and neurophysiological domain, to estimate affective food experience.

A challenge in using implicit measures to estimate affective food experience, as well as in affective computing in general, is the lack of a ground truth measure of affect (Brouwer et al., 2015c). While we know that implicit measures such as the ones mentioned above are associated with affect, their relation is not clear-cut and context dependent. Therefore, when modelling affect using (combinations of) implicit measures, we need ground truth affective labels as determined within the context of interest. These labels are commonly obtained by ratings of affective experience as given by the individual whose experience is modelled, or as given by other individuals (e.g. when using affective stimuli from standardized stimuli collections such as the affective picture collection IAPS (Lang et al., 2005)). However, as argued in the first paragraph, these ratings have certain properties (such as their conscious and possibly biased nature) that we want to avoid using implicit measures. We should not base the interpretation of implicit measures on exactly the measures whose properties we would like to avoid.

In the current study, we therefore follow an approach that, in principle, does not require precise affective ratings. In this approach we train a model using the vinegar solution as an extreme high arousal, low valence stimulus, and then apply the model to the regular drinks. The vinegar solution may turn out to be an outlier that cannot be used to generalize to regular drinks. As described earlier, for some measures, this may be the case. However, if it does generalize sufficiently (as evidenced by an association between the model's output score and self-reported ratings for regular drinks), this approach would be a promising way to gauge individuals' affective experience when tasting food without needing to ask. As a comparison, we contrast this approach to a more traditional approach that does use the valence and arousal ratings as the 'true' affective experience of drinks.

Methods

Participants

A total of 70 healthy participants (19 men, 51 women) took part in this study. They were between 19 and 63 years old, with an average of 48.5 years. All participants signed an informed consent in accordance with the Helsinki Declaration of 1975 as revised in 2014 (WorldMedicalAssociation, 2014), before participating in this study. The study was approved by the TNO Institutional Review Board (TCPE). Three participants were excluded due to technical problems related to the registration of event markers and physiological data recording. This left us with 67 participants for further analysis.

Materials

EEG, ECG and EDA were recorded using a Biosemi Active Two MkII system, with a sampling frequency of 512 Hz. For EEG, 32 active silver-chloride electrodes were placed according to the 10-20 system. ECG electrodes were placed on the right clavicle and on the lowest floating left rib. EDA was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. Pupil diameter was recorded at 60 frames per second using SmartEyePro V6.1.6. This system consists of two cameras, placed at the left and right side of a screen that presented the name of the drink and the rating scales. The screen had a size of 37.0 by 30.0 cm and the viewing distance was approximately 80.0 cm. Room lightening was normal office lightening, and the same for all participants.

SAM pictures (Bradley and Lang, 1994) with 9-point scales were used for valence and arousal self-report ratings. The 9-point scale was positioned in the appropriate location at the bottom of each SAM scale, where the most leftward (most unpleasant and calm) and the most rightward (most pleasant and aroused) parts of the scale were translated into values of 1 and 9, respectively. Participants entered their rating by clicking on one of nine small circles.

For the behavioral measure of sip size, the exact weight of each drink including the cup was measured before the participant took a sip. After finishing the experiment, the cups with the rest of each drink were weighted again to determine the sip size in grams.

The drinks used in this study were apple juice, orange juice, yogurt drink, milk, buttermilk, rooibos tea, black tea, cola and a vinegar solution (50% vinegar, 50% water). The regular drinks were chosen to represent a variation in basic flavors and temperature. They clearly differed from one another in taste, but they were expected to be close to one another in affective experience, at least relative to diluted vinegar. Teas and the vinegar solution were prepared in the same way each morning. Teas were kept at about 60 °C, and the vinegar solution was kept at room temperature. The other drinks were kept in a refrigerator before being served to the participants. Each sample was served in a white plain cup, in portions of 50 g. Participants tasted the drinks in randomized order, except that the vinegar solution was always followed by a cup of water, to decrease possible lingering of emotional and physiological effects. Responses to water are therefore not included in the analyses.

Design and Procedure

After participants arrived at the laboratory, the experimental procedure was explained, and they were asked to sign the informed consent. The electrodes for EDA, ECG, and EEG were attached, and participants were asked to sit comfortably in front of a screen. Participants filled out a general questionnaire on demographic de-tails and current emotional state. Before the experiment started, the SmartEyePro system was calibrated. Then, the experimenter showed and explained how to take a sip, immediately putting the cup down after the sip, and participants performed a practice trial with water. After this there was time for additional practice or instructions when needed, and participants had the chance to ask questions. The timeline of an experimental trial is indicated in Figure 4-1. First, the name of drink was presented on the screen. This was the sign for the experimenter to place the appropriate drink in front of the participant. After 5 seconds, the name of the drink disappeared, which was the sign for the participant to take one sip. Participants were instructed and practiced to always make the same drinking movement. After taking the sip, participants sat still and looked at a blank white screen. Forty seconds after the name of the drink had appeared on the screen, the self-report valence and arousal rating scales appeared. After rating valence and arousal, the name of next drink appeared on the screen. This procedure was repeated until all ten drinks (eight regular drinks, vinegar solution and water) had been served once. Note that in our procedure, and as common in daily life, participants always knew what drink they were about to taste. The food experience and physiological processes that we study thus reflect a mixture of expectation and sensory processes, starting at the moment that the name of the drink appears on the screen.

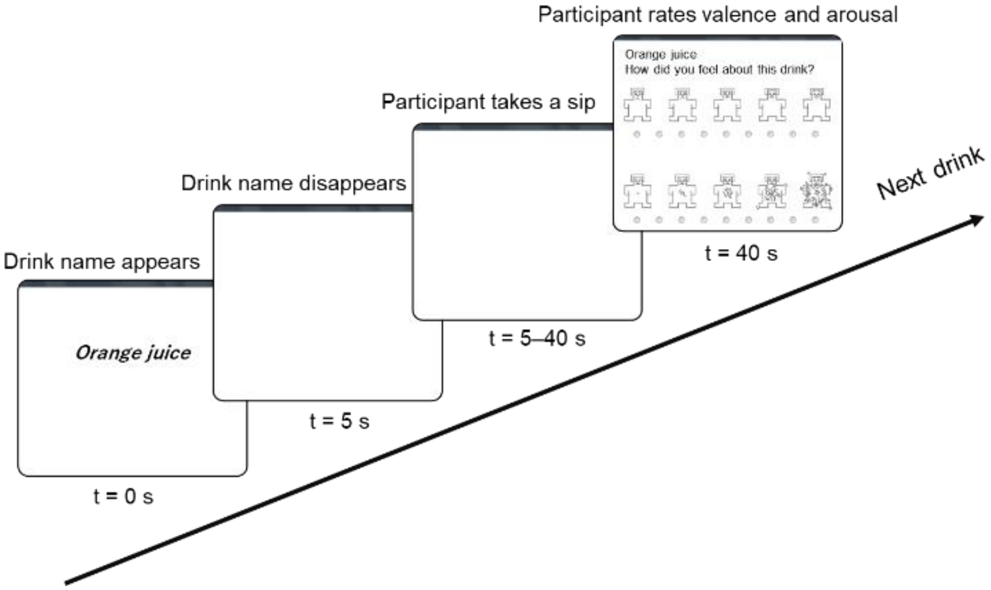


Figure. 4-1. Outline of an experimental trial. After participants rated the drink, the next trial started in which another drink was presented. In this manner, each participant responded to each of the ten different drinks once.

Data Processing and Analysis

Before further analysis, EDA, ECG and EEG were preprocessed.

The EDA signal was bandpass filtered between 0.03 and 100 Hz with custom made Matlab algorithms.

Inter-beat interval (IBI), defined as the temporal distance between R-spikes, was extracted from the ECG signal using custom made Matlab algorithms.

Raw EEG data were pre-processed and analyzed using MATLAB and the FieldTrip open source MATLAB toolbox (Oostenveld et al., 2011). The EEG pre-processing entailed standard procedures of referencing the signals to the average EEG signal and filtering them using a 0.5 Hz high pass and a 43 Hz low pass filter to remove slow drifts and high-frequency noise. Logistic infomax independent component analysis (ICA, (Bell and Sejnowski, 1995)) was performed to classify artifactual independent components, i.e., components not reflecting sources of neural activity, but rather ocular or muscle-related artifacts. These components were re-moved from the data. This was done using EEGLAB v14.1.2 for MATLAB (Delorme and Makeig, 2004).

Valence and arousal ratings were normalized by subtracting the average rating of a participant of each of her/his ratings so that for each participant, the average valence and arousal rating was 0.

Table 4-1 gives an overview of all extracted features. Summing the number of features gives a feature vector length of 152 for each drink and each participant. The features were extracted as described in the following.

Sip size was log transformed to obtain more uniform distributions and normalized like the valence and arousal ratings.

The EDA signal was cut in 40s epochs, starting at the onset of the name of the drink and ending 40 seconds after, for each drink and each participant. Using a Matlab toolbox for analyzing EDA (Ledalab, (Benedek and Kaernbach, 2010)), several features reflecting phasic and tonic activity were extracted from the raw signal, where possible using both Continuous Decomposition Analysis (CDA) and Through To Peak (TTP) analysis. Values were normal-ized as described above for valence and arousal.

Several time series features were determined. For this, pupil size, EDA, IBIs and EEG were cut in eight 5s epochs, starting at the onset of the name of the drink and ending 40 seconds later, at the time that the participants scored their experience. For each of these 5s epochs, features were extracted (e.g., the mean or median over these 5 seconds) to form a time series response. Note that 5s intervals are too short to quantify parasympathetic activity from heart rate variability (Camm et al., 1996). Heart rate variability as computed here should thus be seen as an indicator of the extent to which successive IBIs change.

To arrive at EEG alpha asymmetry, the following steps were taken. For each of the intervals, the spectral power was calculated over bands ranging from 8 to 13 Hz (alpha) in steps of 0.2 Hz following a fast Fourier transform (FFT) approach using a single Hanning taper. Subsequently, values were integrated. FAA at F3 (F7) and F4 (F8) was determined for each 5s segment by taking the relative difference between alpha as recorded at the right and the left side of the cortex: ((R - L) / (R + L)) *100 (Papousek et al., 2014). Positive values indicate lower alpha power in the left than in the right hemisphere (i.e., relatively greater left hemisphere cortical activity). Besides alpha asymmetry, alpha at Pz and theta (4-8 Hz) at Fz were included as variables, since these might be informative of agitation or arousal (Pfurtscheller et al., 1996; Bekkedal et al., 2011).

All time series variables, except for heart rate variability, were normalized per individual, drink and variable, by subtracting the series means from the series. In addition to the 8 values reflecting the subsequent 5s epochs, a 9th value reflecting the mean was added to form time series of nine values.

For each variable (feature), before and after normalization, values that were more than 5 standard deviations away from the mean were discarded as outliers. These missing data (<2%) were imputed using k-nearest neighbors imputation.

Table 4-1. An overview of all extracted features

Sensor category	**Main variable**	**Variable version**	**Nr. of features**
Behavior	Sip size	Sip size	1
Pupil diameter	Pupil diameter time series	Median	9
		Standard deviation	9
EDA	Raw skin conductance level time series	Mean	9
		Median	9
		Minimum and maximum of mean	2
		Minimum and maximum of median	2
	Phasic skin conductance (Ledalab)	Number of responses (CDA and TTP)	2
		Summed response amplitude (CDA and TTP)	2
		Response latency (CDA and TTP)	2
		Global mean	1
		Global maximum deflection	1
		SCR: average response	1
		Phasic maximum	1
		Time integral of response	1
	Tonic skin conductance (Ledalab)	Tonic	1
ECG	Inter beat interval time series	Mean	9
		Median	9
	Heart rate variability time series	RMSSD	9
EEG	Alpha asymmetry time series	F3F4	9
		F7F8	9
	Alpha time series	F3	9
		F4	9
		F7	9
		F8	9
		Pz	9
	Theta time series	Fz	9

Modelling

For our approach to use the vinegar solution as ground truth high arousal, low valence; and the regular drinks as ground truth low arousal, high valence stimuli, we trained an elastic net binomial classification model to distinguish between vinegar and all other drinks. All implicit variables were used for this. The model used 10-fold cross validation to determine the optimal lambda and to obtain error rates. More specifically, the procedure ran for 11 (10-folds+1) times; the first time to get the lambda sequence, and then the remainder to compute the fit with each of the folds omit-ted. The average error and standard deviation over the folds were computed. Minimum average classification error was reached for a certain lambda which led to the final classification accuracy. Up-sampling (random sampling with replacement) was used to deal with class imbalance (i.e., the fact that vinegar was a less frequent stimulus compared to regular drinks). The final error rate is reported for balanced classes.

In addition, combinations of variables were tested with the idea of choosing a limited, but still sensitive sensor set. For this, we tested all pairs of the ‘sensor categories’ behavior, pupil diameter, EDA, ECG and EEG.

For the comparison approach (using explicit valence and arousal as ground truth labels), an elastic net regression model was used using all implicit variables. Again, 10-fold cross validation was used to find the optimal lambda and to obtain error rates. For this approach, only data from regular drinks was used.

Results

Approach 1: Using the Vinegar Solution as Ground Truth

Training a linear elastic net binomial classification model on all implicit variables using balanced classes ‘vinegar’ and ‘regular drinks’ resulted in a classification accuracy of 91%. The trained classification model was then used to assign continuous scores to the regular drinks. A high score means that the data is judged to be closer to vinegar (closer to high arousal, low valence); a lower score means that the data is judged closer to regular drinks (closer to low arousal, high valence). Indeed, we found a positive correlation between the mean drink scores as assigned by the model, and mean rated arousal (Pearson’s *r*, $r = .64$, $p = .04$, Figure 4-2A). We also found a trend of a positive, rather than negative, correlation between the mean predicted scores and mean rated valence ($r = .63$, $p = .05$, Figure 4-2B). As a reference, the model’s assigned score for vinegar was 3.6. Rated arousal for vinegar was 1.0, and rated valence -2.4.

The best combination of any two implicit sensor categories was ECG and sip size (which was in accordance with our finding that the top 23 largest weights in the general model were ECG derived features; sip size ranked 43th of the 152 features). Training the model using features from only ECG and sip size resulted in a classification accuracy of 76% distinguishing vinegar from regular drinks. Similar as what we found when using the model above, mean drink scores as predicted by this model correlated positively with mean rated arousal ($r = .87$, $p < .01$, Figure 4-3A). No correlation with valence was found ($r = .46$, $p = .13$, Fig 4-3B). The score of vinegar in this model was 0.8.

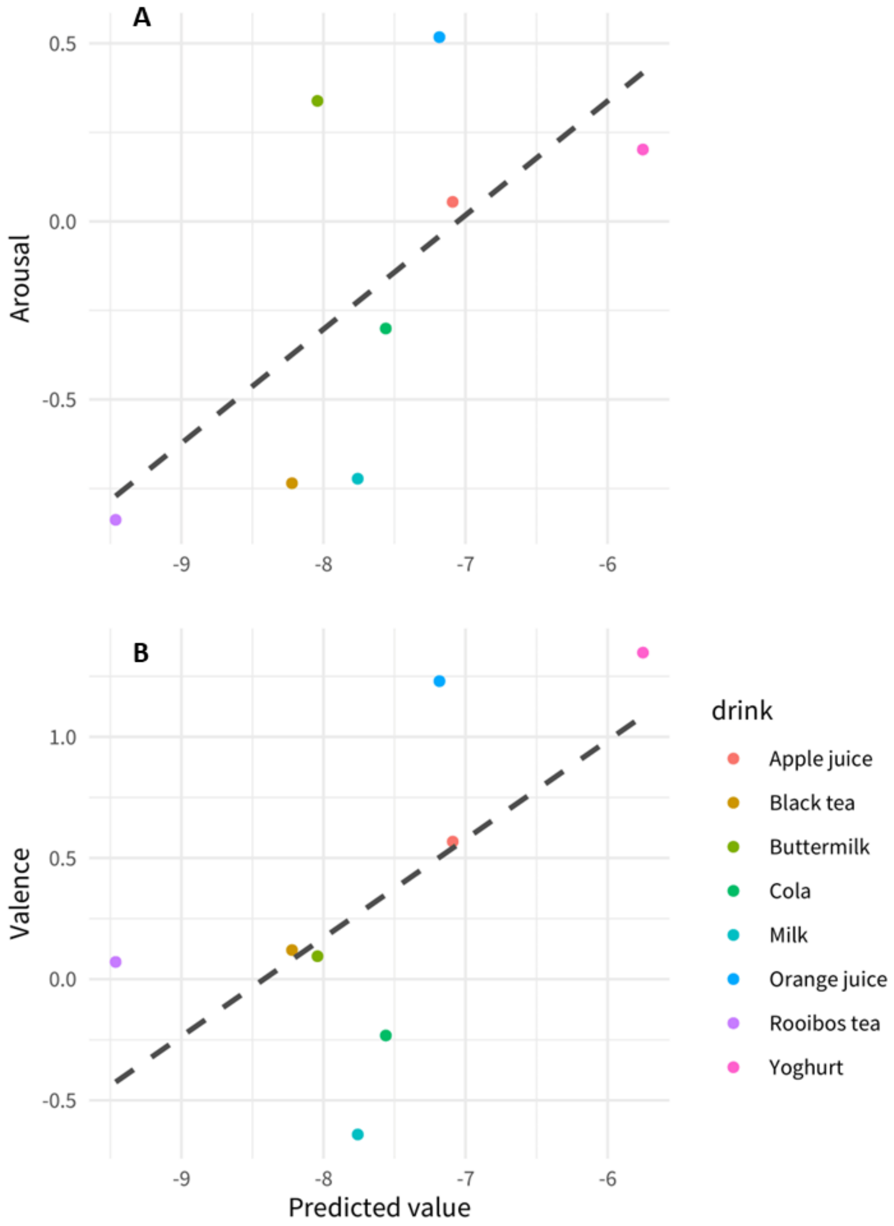


Figure 4-2. Scores assigned by the model trained on all implicit features versus rated arousal (A) and rated valence (B), averaged across drinks.

Approach 2: Using Explicit Valence and Arousal as Ground Truth

Training an elastic net regression model using all implicit variables to predict arousal ratings and valence ratings for all regular drinks was not successful. Model performance as judged by RMSE values obtained during the cross-validation procedure was very low. The model reached best performance (i.e. lowest RMSE values) with variable coefficients of 0; this means that any linear combination of predictors will result in worse performance compared to simply using the mean rated arousal or valence of all drinks as the predicted value.

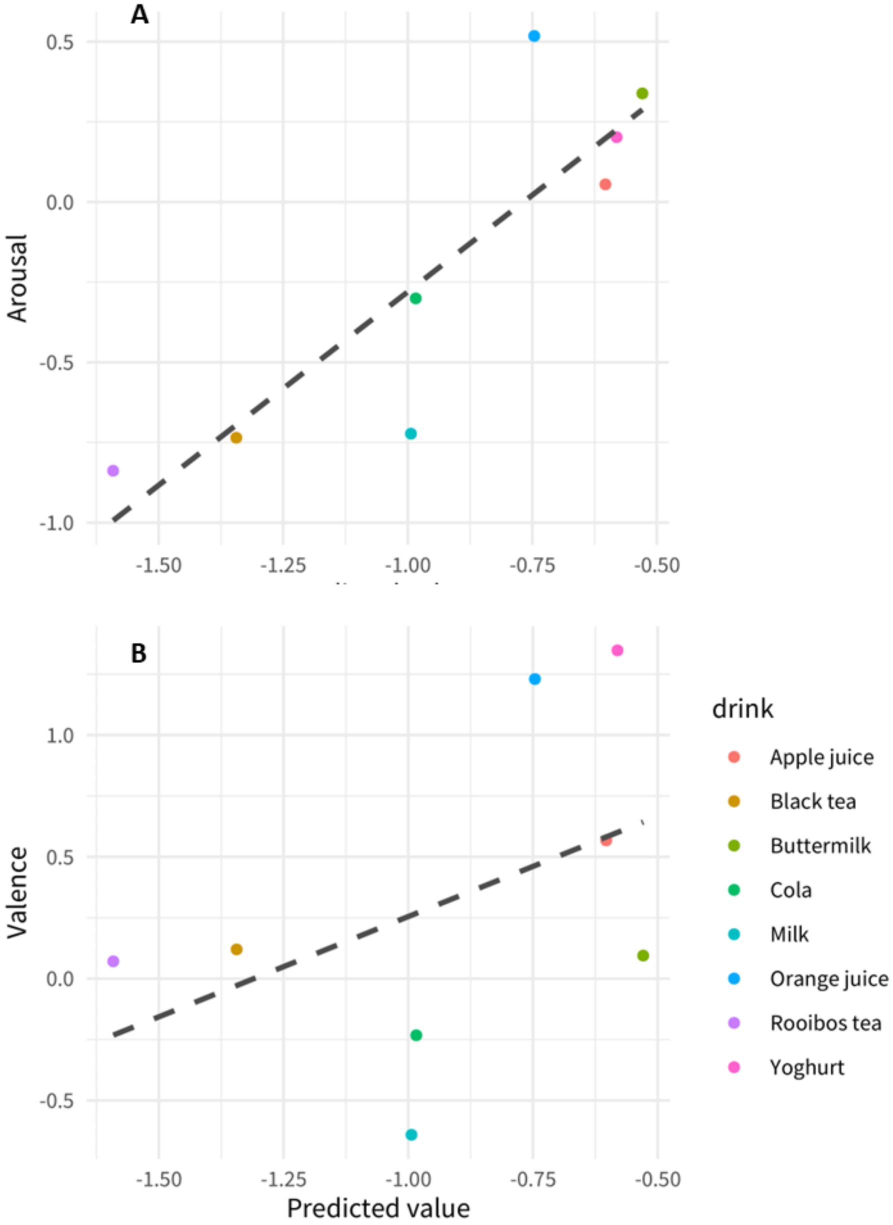


Figure 4-3. Scores assigned by the model trained on features derived from ECG and sip size versus rated arousal (A) and rated valence (B), averaged across drinks.

Discussion

In the current study, we explored the potential of combining implicit measurements, covering the behavioral, physiological and neurophysiological domain, to estimate affective experience when tasting drinks.

Training a model using sip size, pupil diameter, ECG, EDA, and EEG using the vinegar and regular drinks data performed well in attributing unseen data to either of these two categories (91% correct). Thus, it is possible to estimate from implicit data from an individual tasting a particular drink, whether this individual tasted a regular drink or an unpleasant, high arousal drink. When limiting the number of sensor categories to two, ECG and sip size are the best choice, reaching a classification accuracy of 76%. Relatively high sensitivity of heart rate and sip size to distinguish vinegar from regular drinks is consistent with our earlier analysis of these data (Kaneko et al., 2019a).

For most potential applications, models should be able to estimate affective experience for stimuli that differ more subtly in affect than the vinegar solution and regular drinks. When we only used the implicit data obtained from individuals tasting regular drinks, and used rated arousal and valence as ground truth affective experience, it appeared to be impossible to estimate rated arousal and valence from unseen data. However, when we used the model described above, trained by only using implicit data obtained from individuals tasting regular drinks as well as a known ground truth high arousal low valence drink (vinegar), and applied this model to regular drinks, model scores correlated with rated arousal in the direction as expected. For valence, there was an unexpected positive correlation trend, where a high score (‘similar to vinegar’) tended to correspond to high rather than low valence. As discussed and substantiated in (Kaneko et al., 2019a), we think that this is due to the fact that in regular drinks, rated valence and arousal correlate positively. Most of our implicit measures reflect arousal, such that the model distinguishing between vinegar and regular drinks capitalizes on the difference in arousal rather than valence. Together this leads to the observed positive trend in the correlation between predicted score and valence in regular drinks. Our results indicate that implicit responses to a quite extreme, affective ground truth stimulus can be generalized to responses to stimuli that differ more subtly in affective experience in the case of tasting drinks. Starting with affective ratings as ground truth (as is usually done) did not work out in this case, possibly because the ratings were too variable and noisy. Note that, in order to validate our approach, we did make use of affective ratings (showing the correlation between rating and model score), but in principle, this is not required.

As found before, our results indicated that implicit measures generally better match arousal than valence. Most (food related) research focusses on measures reflecting pleasantness. At first sight, this may seem the more interesting measure, but if one is interested in stimuli of which it is a priori known that they are all relatively pleasant (as is the case with most food stimuli), arousal may be of special interest. This is especially so since arousal affects the memorability of an event (Anderson et al., 2006; McGaugh, 2006). In Köster and Mojet (2006) and Köster and Mojet (2007) arousal is discussed as a crucial determinant in defining (sustained) attractiveness of products. In addition, explicit measures of arousal have been found to be vulnerable to the exact formulation of the question (Toet et al., 2018a), which renders implicit measures all the more useful.

One may have expected that for arousal, EDA would have turned up as the most important predictor. Sweat glands are exclusively innervated by the sympathetic branch of the Autonomic Nervous System and EDA (Dawson et al., 2017) has been associated consistently with arousal across a range of contexts, whereas the link between heart rate and arousal is less clear. While associations between heart rate and arousal are usually positive, negative associations have been found as well (e.g. within the context of reading a novel (Brouwer et al., 2015b)). However, working against EDA is the large variability between individuals (e.g. (Thammasan et al., 2020)), which may have resulted in our finding that within the context of tasting drinks, ECG related variables are more suitable to estimate arousal than EDA.

As explained in the introduction, we chose for a design resembling real-life, natural tasting of drinks. This resulted in certain limitations, such as added noise and being unable to disentangle expectation and sensory effects on food experience and physiological processes. We also do not know to what extent the pure effect of chemical properties of the drinks (or more precisely, their interaction with the senses of an individual) are responsible for the results. For instance, perceived sourness is one of the properties that differs between vinegar and the regular drinks, and the model may have captured that instead of arousal. If arousal and perceived sourness are associated in regular drinks, this could explain the association between rated arousal and the model's score. To get an impression of the importance of this possibility, we trained a linear model to use implicit measures simply for distinguishing between the different drinks. If implicit measures respond relatively strongly to more basic, chemical properties rather than affect, this should be possible. However, this model's classification accuracy (5%) did not exceed chance level (which is 12.5%), suggesting that differences in chemical composition of our drinks do not result in clear differences in implicit responses.

For future research, including multiple types of ground truth stimuli would be valuable. Including a stimulus that is a priori (i.e., for everyone) known to be high arousal - high valence relative to the regular stimuli would be good, but in the case of food stimuli hard to do because regular food stimuli are all high in valence. Still, including multiple types of high arousal - low valence stimuli (e.g. a quinine solution next to the vine-gar solution in the present study) is possible and may result in better generalizability and robustness. While this study showed that we already can generalize a model trained on distinguishing a vinegar solution from regular drinks to estimating affective experience in regular drinks, this was shown only for rated arousal and not for rated valence. Also, the finding that the model that distinguished the two classes best (91% ac-curacy for the model using all features versus 76% for the model using ECG and sip size) did not result in the highest correlation between model scores and rated arousal for regular drinks (cf. Fig 4-2A and 4-3A), indicates that generalizability was not perfect. This may improve when multiple instances of ground truth stimuli are included.

Conclusion

We demonstrated that food experience (arousal) elicited by a range of regular drinks could be estimated using implicit measures without first having to rely on the measures that we aim to avoid, namely, ratings of affective experience. Estimates of affective experience purely based on implicit, 'unconscious' responses may be helpful to understand consumers' food-evoked emotions better compared to only using self-reports. Future research should focus on improving the model e.g. by using different ground truth stimuli, and on relating scores resulting from models based on implicit measures, as well as explicit measures, to variables of interest such as future food choice behavior.

Acknowledgment

The authors wish to thank Esther Bosch, Wouter Korteling, and Benjamin de Graaff for their assistance with conducting the experiment. This work was supported by Kikkoman Europe R&D Laboratory B.V.

Chapter 5

Emotional State During Tasting Affects Emotional Experience Differently and Robustly for Novel and Familiar Foods

This chapter is published as:
Kaneko, D., Brouwer, A.M., Hogervorst, M., Toet, A., Kallen, V., and Van Erp, J.B.F. (2020). Emotional state during tasting affects emotional experience differently and robustly for novel and familiar foods. Frontiers in psychology 11, 2372.

Abstract

Emotional state during food consumption is expected to affect food pleasantness. We hypothesize that a negative emotional state reduces food pleasantness, and more so for novel foods than for familiar foods because novel foods have not yet been associated with previous emotions. Furthermore, we expect this effect to be stronger when judging the food again from memory without tasting. We induced a positive emotional state in 34 participants by telling them that they earned a monetary bonus, and induced a negative emotional state in 35 other participants by subjecting them to a social stress test. After this emotion induction, both groups tasted and rated a (for them) novel soup (sumashi soup) and a familiar soup (vegetable soup). Several explicit and implicit measures of food pleasantness (rated valence, EsSense25, willingness-to-take-home and sip size) indicated that while the negative emotion group did not experience the soups as less pleasant than the positive emotion group, there was an interaction between food familiarity and emotional group. The positive emotion group experienced novel and familiar soups as equally pleasant, while the negative emotion group experienced the novel soup as relatively unpleasant and the familiar soup as pleasant. The latter result is consistent with a comforting effect of a familiar taste in a stressful situation. This effect remained in the ratings given one week later based on memory, and even after re-tasting. Our results show that emotional state affects food pleasantness differently for novel and familiar foods and that such an effect can be robust.

Introduction

Food judgements (as probed by e.g. ratings of food preference or liking, food pleasantness, food choice, and eating behavior) depend not only on the quality of the taste but also on the emotional state during food consumption, the social-emotional context in which the food is consumed and already existing associations between food and emotion (Desmet and Schifferstein, 2008; Salvy et al., 2008). Food associations are related to regional food habits, different food cultures, and food traditions in the family.

The effect of ambience on food intake and food choice was reviewed by Stroebele and De Castro (2004). They define ambience as a context of environmental stimuli and conclude that there are major influences of ambience on eating behavior. The studies that they reviewed showed effects of social-emotional aspects of context, as well as effects of physical aspects of contexts (e.g., colors (Spence et al., 2010; Chen et al., 2018a), sounds (Spence and Shankar, 2010; Woods et al., 2011), and/or odors (Herz et al., 2004)). We assume that many of the reported context effects, especially the social-emotional context effects, influence food judgements through the induction of a certain emotional state. Below we review studies that examine the effect of emotional state on food experience in some more detail.

Birch et al. (1980) examined the effects of pairing positive experiences with snack foods on children's liking of the foods. In their study, the same snack foods were served to children 1) as a reward, 2) by a friendly adult, 3) in a nonsocial context, or 4) at normal snack time. Children's liking ratings were higher on snack foods in the two emotionally positive contexts (as a reward or by a friendly adult) than in the other contexts, indicating that the liking of snack foods was affected by emotion. Siegel and Risvik (1987) examined the effect of positive and negative mood on acceptance ratings of an almond dairy bar in adulthood. They induced different moods by asking participants to indicate their current state using questionnaires that either contained positively formulated statements such as "I feel great" (positive mood group) or negative statements such as "I feel weak" (negative mood group). They found that participants from the positive mood group reported significantly higher acceptance of the almond dairy bar than those in the negative group. Kuenzel et al. (2011) aimed to induce different emotional states using video clips to investigate the effect of emotional state on food preference and liking. For five consecutive days, participants watched 4 to 5 minutes positive video clips (two different positive states: one active and one relaxed) or a neutral video clip. Two different novel uncolored drinks were developed for this study: a generally liked drink and a more neutral drink. Participants were served the liked or the neutral drinks just before the start of the film clips and were instructed to finish them by the end of the clips. The study showed an interaction between type of drink (neutral and liked) and emotional state (active, relaxed and neutral) indicating that liking ratings of the liked drink were lower in the relaxed condition than in the neutral condition. Thus, this study shows a (modest) effect of emotional state affecting liking scores of flavored drinks, be it not simply in the direction of an emotional state that was intended to be positive leading to higher liking. The authors suggest that the reason for this effect might be that participants' attention may have been divided when being in a positive state, leading to a tendency to score towards the middle of the liking scales, which resulted in relatively low liking scores for the liked drink.

Walsh and Kiviniemi (2014) used an ‘implicit priming paradigm’ to create one of three emotional associations (positive, negative, or neutral) to images of fruit. This paradigm involves repeated presentation of sequential pairs of a positive, negative, or neutral image or word followed by an image of fruit. Twenty of these pairs were interspersed among a total of 230 images that were presented to each participant. At the end of the experiment, participants were asked to choose one among a selection of apples, bananas, and granola bars. Those in the positive condition were more likely to select fruit compared to those in either the neutral or negative condition.

All the studies discussed above indicate that emotional state can affect experienced food pleasantness and liking. They all used familiar foods as stimuli, except for the study by Kuenzel et al. (2011), in which only unfamiliar stimuli were used. We think that food familiarity is a key factor that may interact with emotional state when experiencing and judging food. When tasting a food for the first time (a novel food), effects of emotional state may be more pronounced than when tasting a familiar food since there is no influence yet of existing associations. Knowledge about such effects in the absence of prior existing associations is important, for instance, when introducing new products to the market, or in medical settings where patients need to consume specific foods, supplements, or medicines. However, we are not aware of research exploring whether the effect of emotional state on food pleasantness indeed differs between novel and familiar food.

In addition, the majority of studies on the effect of emotional state or context focus on the instantaneous effect on food pleasantness. However, Köster and Mojet (2015b) argued that the role of memory is probably much more important than the ‘first impression’ experience that is commonly investigated. They emphasized that products should be tested for the emotions they evoke before, during, a few hours after, and a week (or even longer) after consumption, to obtain a more complete picture of the experience of the product.

In the current study, we evaluate how novel and familiar foods (two types of broths, from now on referred to as soups) are affected by emotional state during tasting (positive/negative), both instantaneously and a week later. We asked participants to come to the lab twice, separated by an interval of one week - we refer to the first day as Day 1 and to the second day a week later as Day 2. On Day 1, participants were asked to taste and rate a novel soup, and a familiar soup. Before tasting and rating the soups, we induced a positive emotional state in half of the participants and we induced a negative emotional state in the other half. On Day 2, participants underwent two separate sessions. In the first session, participants were asked to rate the same soups as tasted and judged on Day 1, but without tasting (i.e. from memory). In the second session, they rated the same soups again, but this time with actual tasting. The effect of emotional state on food experience of novel and familiar soups was not only measured by using self-report (valence ratings; EsSense25 questionnaire (Nestrud et al., 2016b) that probes 25 emotions associated with food), but also by using behavioral measures, namely sip size and willingness-to-take-home. These measures are of a more implicit nature, and expected to support the self-report of valence ratings (Lagast et al., 2017; Kaneko et al., 2018a).

The following hypotheses are tested:

1) Overall experienced food pleasantness, as reflected in the valence ratings and the EsSense25, is lower when tasting soups in a negative emotional state than in a positive emotional state.

2) This effect is stronger for the novel soup than for the familiar soup.

3) Differential effects of emotional state on novel and familiar soups will be stronger a week later when the actual taste of the soup is not available. This is because measures of experience will then only be based on memory, where the novel soup has only been associated with the experience of the (emotional) tasting session in the lab, and the familiar soup is also associated with other, previous food experiences.

4) When participants subsequently taste the soups again, the effect mentioned under 3 is reduced, since experience is no longer based on memory alone.

5) The behavioral measures of sip size and willingness-to-take-home show a similar pattern of results compared to subjective ratings.

In sum, this study will inform us about the interaction between emotional state and food familiarity on food experience, both during initial tasting and a week later.

Materials and Methods

Participants

A total of 70 healthy participants (19 men, 51 women) was recruited for this study. Exclusion criteria were food allergies or special diets. One of the male participants dropped out from the study. Data from this participant were excluded from all analyses, leaving us with data of 69 participants. All participants had the Dutch nationality and were between 19 and 63 years old, with an average of 48.4 years and a standard deviation of 10.4 years. Participants were recruited through the participant pool of the research institute where the study took place (TNO) and received a basic monetary reward of 30 Euro per participant to compensate for time and travel costs. On top hereof, and unknown to them beforehand, participants in the positive emotional state group received a 5 Euro bonus. Before participating in this study, all participants signed an informed consent in accordance with the Helsinki Declaration of 1975 as revised in 2014 (WorldMedicalAssociation, 2014). The study was approved by the TNO Institutional Review Board (TCPE). After signing the informed consent, they were randomly assigned to the positive emotional state group (34 participants: 10 men, 24 women, average age of 49.2 years) or the negative emotional state group (35 participants: 8 men, 27 women, average age of 47.7).

Sip size data was not complete for three participants (one from the positive group, two from the negative group), and were thus left out in the analysis on sip size.

We also recorded physiological data. These recordings failed for two participants (one from the positive group, and one from the negative group), and were thus left out in the analyses on the physiological data.

Materials

Test stimuli

Vegetable and sumashi soup were selected as familiar and novel soups, respectively. Vegetable soup was prepared using vegetable bouillon cubes (Maggi, Nestle, Switzerland) following the instruction on the package. Sumashi soup is a traditional Japanese transparent soup. It was prepared by mixing 5.0 g of seaweed broth (Riken Vitamin, Japan), 20.0 g of soy sauce (Kikkoman, Japan), 5.0 g of cooking sake (Wadakan, Japan), and 1.0 g of sea salt with 750 ml of hot water. The two soups were always prepared in the same way each morning and kept at approximately 60 ^{o}C until they were served. Before serving the soups, a selection of regular drinks (apple juice, orange juice, yogurt drink, milk, buttermilk, rooibos tea, black tea, cola), diluted vinegar (50% vinegar, 50% water) and water were served in semi-randomized order. This was done to answer other research questions (Kaneko et al., 2019a). All soups were served in white plain cups, in portions of 50 g. At the end of Day 2 of the experiment, 100 g of each of the two soups was given to further assess the emotions evoked by tasting each soup.

Valence scale

SAM pictures (Bradley and Lang, 1994) with nine-point scales were used for valence self-report ratings. The nine-point scale was positioned in the appropriate location at the bottom of each SAM scale, where the most leftward (most unpleasant) and the most rightward (most pleasant) parts of the scale were translated into values of 1 and 9, respectively. With respect to valence, participants were asked how pleasant their experience with the soup was, with the manikin on the right indicating a very pleasant experience and the manikin on the left a very unpleasant experience. Participants were instructed that they should try to answer quickly, without thinking too long.

EsSense25 questionnaire

Besides valence scales, the EsSense25 (Nestrud et al., 2016b) was used to obtain self-reported emotions evoked by experiencing the two soups. The EsSense25 is a shorter version of the EsSense Profile® (King and Meiselman, 2010), which was developed to measure emotions associated with foods. Each of 25 emotional terms (*Loving, Nostalgic, Good, Good natured, Joyful, Bored, Secure, Happy, Warm, Disgusted, Pleasant, Active, Satisfied, Aggressive, Guilty, Calm, Free, Understanding, Enthusiastic, Interested, Tame, Adventurous, Wild, Mild, and Worried*) was assessed on a five-point scale ranging from 1 (not at all) to 5 (very much).

Behavioral measures

For the behavioral measures, sip size and willingness-to-take-home were recorded. To measure sip size, the exact weight of each soup including the cup was measured before the participant took a sip. After finishing the experiment, the cups with the remaining of each soup were weighed again to determine the sip size.

A modified rating scale of willingness-to-take-home (Wichchukit and O'Mahony, 2010b) was used in this study. While in the original scale, participants would be asked which soup as used in the experiment they wanted to take home as a reward, we asked participants how many cups of each soup they would want to take home after the experiment with a maximum number of 6 cups (e.g., 1 sumashi soup and 5 vegetable soup). Participants could choose less than 6 cups in total (e.g., none, or 2 sumashi soups and 3 vegetable soups).

Physiological recording equipment (EDA and ECG)

EDA (electrodermal activity - for skin conductance level or SCL) and ECG (electro cardiogram, for inter-beat interval or IBI which is the inverse heart rate) were measured to assess whether the experimental induction of emotion was effective in case we would not find any effect of emotion. EDA and ECG were recorded using an Active Two MkII system (Biosemi B.V., Amsterdam, The Netherlands), with a sampling frequency of 512 Hz. SCL was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. ECG electrodes were placed on the right clavicle and on the lowest floating left rib. SCL was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. EEG (electroencephalogram) was recorded as well, for different research questions (Kaneko et al., 2019a).

Emotional state induction

On Day 1, participants underwent either one of two types of emotional state induction, depending on the group they were assigned to. To induce a positive emotional state, participants received a message on the screen that they would receive an extra monetary bonus for participating in the experiment, and that after tasting and judging the second soup, they would receive the instruction to flip a card on the table that would tell them the exact amount of this bonus. This message was displayed just before displaying the name of the first soup. After tasting and judging the soups, the message to now flip the card was displayed. Participants flipped the card telling them that the amount of the bonus was 5 Euro. They received this bonus at the end of the experiment. To induce a negative emotional state, we used a modified Sing-a-Song Stress Test which has been shown to induce profound social stress (Brouwer and Hogervorst, 2014; Brouwer et al., 2017b; Toet et al., 2017). Just before the first soup, a message was displayed that they would receive the instruction to sing a song out loud after tasting and judging the second soup. This instruction was given as announced, and participants started singing a song. The aim of these emotion induction procedures was to induce emotions that were as different as possible with respect to pleasantness in the two groups, while keeping other elements (such as receiving an announcement about an exciting task to perform after tasting) as similar as possible.

Experimental Design and Procedure

General procedure Day 1

After participants arrived at the laboratory, the experimental procedure was explained by the experimenter. They were informed that they were going to take part in a study on food experience, in which they would taste and judge drinks and soups. Participants were not told about the emotion induction. After the explanation of the study, participants signed the informed consent and, unknown to them, were randomly assigned to the positive or negative emotional state group. Electrodes for measuring EDA, ECG, and EEG were attached, and participants were asked to sit comfortably in front of a computer screen. Participants were instructed how to take one sip and practiced an experimental trial. At this time, when the participant was in the negative emotional group, one of experimental leaders came in, pretending to be a next participant who arrived at the lab earlier than the appointed time. The other experimental leader asked the fake participant to stay in the same lab room to wait for the previous participant to finish the experiment. Thus, the fake participant was in the room during the whole experiment for the negative emotional state group. In the positive emotional state group, only the experimenter was present. Participants filled out a general questionnaire on demographic details and current emotional state. A tasting and rating trial (schematically depicted in the top left of Figure 5-1) went as follows.

First, the name of the test sample was presented on the screen. This was the sign for the experimenters to place the appropriate cup in front of the participant. After 5 seconds, the name of the test sample disappeared, which was the sign for the participant to take one sip. After taking the sip, participants sat still and looked at a blank white screen. Forty seconds after the name had appeared on the screen, the self-report SAM rating scales appeared. After entering the scores, the next trial started. This procedure was repeated until all drinks had been served (depicted in grey in Figure 5-1). Immediately after, the group-dependent emotional state was evoked through an instruction screen as outlined above. Then, participants in both groups were served the two soups following the same procedure as before, and after rating the second soup, performed the task as instructed (i.e., either flip the reward card or sing a song). Sumashi soup was presented as ‘Asian soup’ to participants. Half of the participants first tasted the vegetable soup, and half the sumashi soup. After participants completed the task, another 100 grams of the two soups were served to all participants in the same serving order as they had tasted and rated before. This time participants could taste more than once and were asked to more elaborately self-report their emotions evoked by tasting each soup using the EsSense25. After filling out the EsSense 25 questionnaires, participants were asked to answer whether they were familiar with the taste of ‘Asian soup’ and to write down the name of the soup if they knew the name or wanted to make a guess. In the end of the experiment, we asked participants how many cups of vegetable soup and Asian soup they would want to bring home if they would receive them for free (with a maximum of 6 in total). They did not actually receive such cups of each soup to prevent them to consume the soups (more than they usually would do) the days preceding Day 2.

General procedure Day 2

The session on Day 2 was divided into two parts and was conducted without any physiological measures and without emotion induction. The schematic experimental procedure is summarized in the bottom half of Figure 5-1. First, participants were asked to sit in front of a screen and rate each drink and soup without tasting them, only relying on their memorized experience from one week ago. The name of the drink or soup appeared on the screen, followed by the SAM scales as on Day 1, but without the 40 seconds blank screen period in between. For each participant, the order of the drinks and soups was the same as on Day 1. Next, participants were asked to rate their emotions with the two soups using the EsSense25, i.e. based only on their memory of the taste and the emotions they had encountered a week before. Then, they were asked again for the two soups they would want to take home in the same manner as on Day 1. After this first session of Day 2 (referred to as Day 2-1), the second session of Day 2 (Day 2-2) commenced. In this second session, the same procedure was repeated, but this time with tasting and rating the drinks and soups in exactly the same manner as on Day 1. This also included judging the soups using the EsSense 25, while being provided with 100 grams of each of the two soups. Finally, as on Day 1, we asked how many cups of vegetable soup and Asian soup they would like to take home. After they completed the experimental tasks in Day 2, the experimenters debriefed participants on the purpose of the study and the emotion induction procedures.

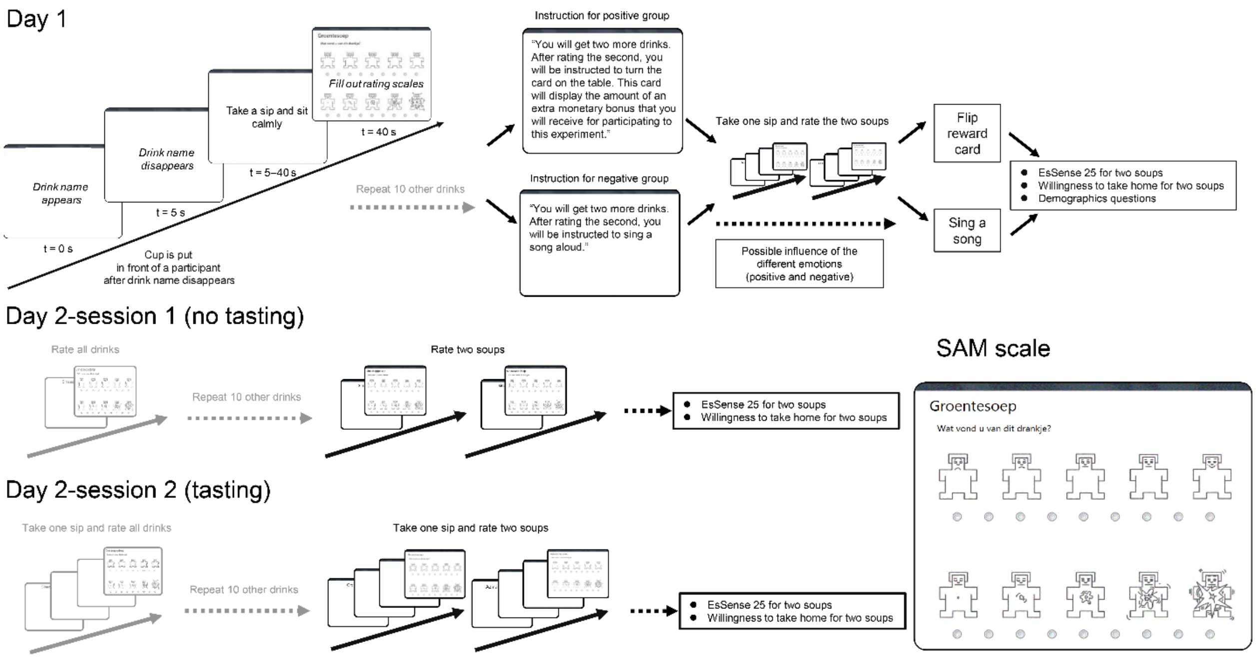


Figure 5-1. Schematic overview of an experimental trial and of the experimental procedure in Day 1, Day 2–1, and Day 2–2. At the bottom-right, a rating screen is depicted, showing the small circles below the SAM scales that participants clicked in order to give their responses, upon which a new trial started. In Day 1, participants are separated into two groups, with either positive or negative emotion induction. Emotion induction occurred before tasting and rating two soups. In Day 2–1 and Day 2–2, all participants followed the exact same procedure. Parts of the procedure highlighted in gray served to answer other research questions (Kaneko et al., 2019a).

Data processing and Statistical analysis

Main dependent variables

Our main dependent variables reflecting food experience are valence ratings, EsSense25 ratings, sip size and willingness-to-take-home.

Statistical analyses on dependent variables were conducted using SPSS ver. 25. To investigate the main effects and interactions of soup (familiar vegetable soup and novel sumashi soup), session (Day 1, Day 2-1, and Day 2-2), and emotional state (positive and negative), we performed mixed model analyses (Maximum Likelihood approach) with soup (2) and session (3) as within-subjects variables, and state (2) as between-subjects variable. For sip size, session involved two rather than three levels (Day 1 and Day 2-2), since participants did not take a sip in the Day 2-1 session.

For all statistical tests, we consider an alpha level of 0.05. Given that the EsSense25 features 25 variables, correction for multiple testing is in place. Therefore we also interpret these results within the light of the Bonferroni corrected alpha level of 0.002. LSD Post-hoc comparisons were performed to interpret any significant interactions that, in the case of EsSense25, survive the Bonferroni correction. This came down to post-hoc comparisons that elucidated state*soup interactions in six measures.

SCL and IBI

Custom made MATLAB 2019a (www.mathworks.com) algorithms were used to extract SCL and IBI. To examine the effect of the instruction that was intended to induce either positive or negative emotion on SCL and IBI, the following steps were followed. First, the EDA signal was bandpass filtered between 0.03 and 100 Hz. Inter-beat interval (IBI), defined as the temporal distance between R-spike (Appelhans and Luecken, 2006), was extracted from the ECG signal using custom made algorithms. Next, for each participant, EDA was averaged across the forty seconds starting at onset of the announcement of the first soup that was presented immediately after the message that induced the positive or negative emotion. The same was done for the last drink that was presented before the message that induced the positive or negative emotion. This latter value served as a baseline. After log transforming the values, the baseline was subtracted from the value obtained after the emotion induction. The same procedure was followed for IBI. An increase in emotional arousal would be reflected by decreased IBI (i.e., increased heart rate) and increased SCL (Brouwer and Hogervorst, 2014). We examined whether these differential values were indeed statistically different from zero using one-sample t-tests. We also compared them between the positive and negative emotional state groups by using two-sample t-test.

Results

Verifying the experimental manipulations

Emotion induction

Figure 5-2 shows the average difference of the mean SCL and the mean IBI before and after the positive emotion induction (announcement to flip a bonus reward card after tasting and rating two soups) and the negative emotion induction (announcement to sing a song after tasting and rating two soups). As expected, IBI decreased and SCL increased for both positive emotion induction (IBI: t (32) = -2.61, p = .014; SCL: t (33) = 2.89, p = .007) and for negative emotion induction (IBI: t (32) = -4.14, p < .001; SCL: t (33) = 6.09, p < .001). This shows that both types of emotion inductions indeed elicited arousal. Two-sample t-tests indicated that elicited emotional arousal was even stronger for the negative emotion induction than for the positive emotion induction (IBI: t (65) = 2.51, p = .015; SCL: t (65) = -3.70, p < .001).

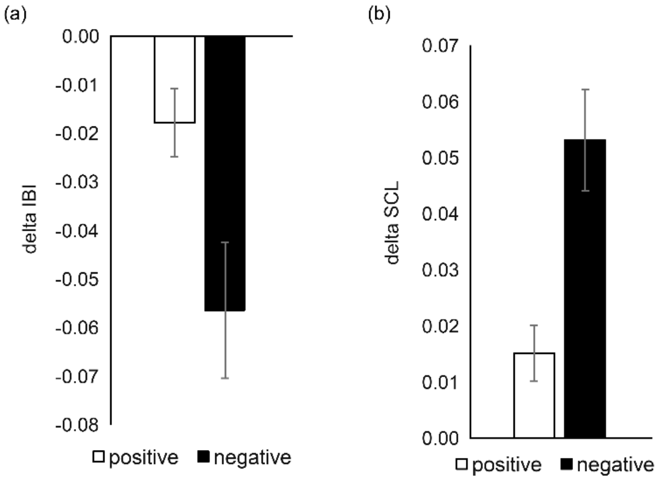


Figure 5-2. Mean delta values for (A) IBI and (B) SCL between before and after announcement in positive emotional group and in negative emotional group. Error bars indicate standard error of the mean.

Novelty of foods

None of the participants reported to have experienced the taste of sumashi soup before and none of them were able to answer the question of what type of ‘Asian soup’ had been used in this study, indicating that the sumashi soup can indeed be considered as a novel soup for all participants in this study.

Effect of emotional state

Table 5-1 presents the results of the mixed model analyses for each of the dependent variables. Significant effects are marked in light grey. Table 5-2 presents post-hoc comparisons that elucidate significant soup*state interactions. In the sections below, we focus on the statistical results that are directly connected to our hypotheses.

Valence ratings

Reported mean valence of each soup averaged across participants for each of the three sessions (Day1, Day2-1 and Day 2-2), each of the two emotional states (positive/negative), and each of the two soups (familiar/novel), is presented in Figure 5-3. There was no main effect of emotional state on valence ratings *(hypothesis 1)* ($F(1, 69) = .09$, $p = .764$, $\eta_p^2 = .001$), but we found a significant interaction effect between emotional states and soups ($F(1, 345) = 20.90$, $p < .001$, $\eta_p^2 = .087$). The post-hoc tests indicated that the novel soup was judged as less pleasant than the familiar soup in the negative emotional state, whereas there was no difference between the ratings of the two soups in the positive emotional state (first two columns in Table 5-2). Post-hoc tests also indicated that the familiar soup was judged as more pleasant in the negative than in the positive emotional state; and that the novel soup was judged as less pleasant in the negative than in the positive emotional state (last two columns in Table 5-2) *(hypothesis 2)*. A lack of interaction between state, soup, and session indicates that this effect remains constant across sessions *(hypotheses 3 and 4)*.

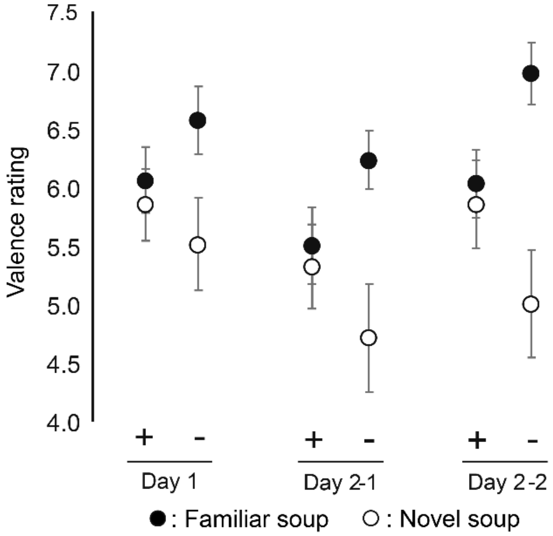


Figure 5-3. Mean valence ratings of familiar and novel soup by positive emotional group (+) and negative emotional group (-) in Day 1, Day 2–1, and Day 2–2. Error bars indicate standard error of the mean.

Self-reported emotions (EsSense25)

Figure 5-4 shows the mean ratings for the 25 emotions of each soup averaged across participants for each of the three sessions (Day1, Day2-1 and Day 2-2) in each of positive and negative emotional state. As can be seen in Table 5-1, similar to what was found for valence ratings, there was no main effect of emotion *(hypothesis 1),* while for nine out of 25 emotions the analyses revealed significant interactions between emotional states and soups *(hypothesis 2)*. Six of these nine emotions concerned positive emotions (*Happy*, $F(1, 345) = 18.80$, $p < .001$, $\eta_p^2 = .109$; *Pleasant*, $F(1, 345) = 11.75$, $p = .001$, $\eta_p^2 = .060$; *Good*, $F(1, 345) = 10.10$, $p = .002$, $\eta_p^2 = .064$;

Warm, *F*(1, 345) = 5.42, *p* = .020, η_p^2 = .043; *Enthusiastic*, *F*(1, 345) = 4.68, *p* = .031, η_p^2 = .023; *Joyful*, *F*(1, 345) = 4.61, *p* = .033, η_p^2 = .027). They all showed the same pattern as rated valence, namely stronger positive emotions for the familiar soup than the novel soup, for negative compared to positive emotional state. Note that only *Happy*, *Pleasant* and *Good* pass the Bonferroni corrected alpha level of *p* = .002. As for valence, post-hoc comparisons indicate that there was no significant difference between the soups in the positive emotional state, but that in the negative emotional state, familiar soup was more positively judged than the novel soup. The negative emotions, *Disgusted* and *Guilty*, that showed a significant interaction between soup and state, revealed a consistent pattern with stronger rated negative emotions for the novel soup than the familiar soup, for negative compared to positive emotional state (*Disgusted*, *F*(1, 345) = 7.56, *p* = .006, η_p^2 = .031; *Guilty*, *F*(1, 345) = 5.82, *p* = .016, η_p^2 = .027). Finally, *Bored* showed a significant soup*state interaction (*F*(1, 345) = 4.54, *p* = .034, η_p^2 = .043), indicating that participants with a positive emotional state rated the novel soup as less boring than the familiar soup for the positive rather than the negative emotional state, fitting with the other EsSense25 and valence results. However, none of the effects found for negative emotions pass the Bonferroni corrected alpha level. For none of the emotions, a significant three-way interaction was found, indicating that the interaction effects between soup and emotional state are stable across sessions *(hypotheses 3 and 4)*.

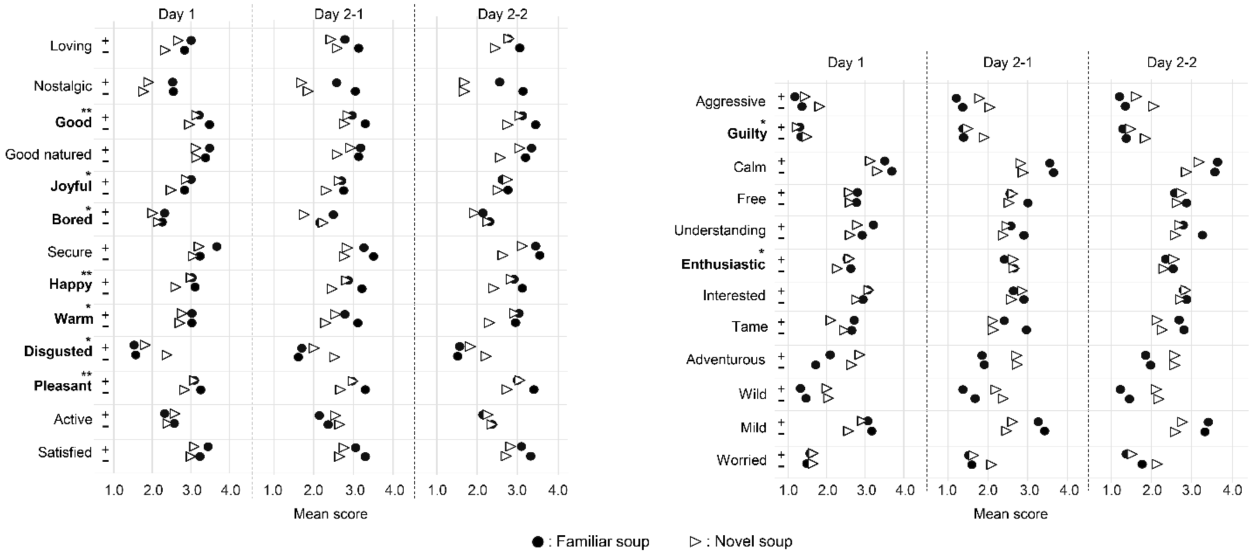


Figure 5-4. Mean rated scores of each emotion based on EsSense25 of familiar and novel soup by positive emotional group (+) and negative emotional group (-) in Day 1, Day 2–1, and Day 2–2. Bolded emotions indicate significant interactions between emotional states and soups with *p < 0.05 and with **Bonferroni correction of p < 0.002.

Behavioral measures: sip size and willingness-to-take-home

Figure 5-5 shows the mean sip size for each soup, each of the two sessions that included sip size (Day1 and Day2-2, not Day 2-1), and each emotional state. Figure 5-6 shows the mean number of cups of soup participants would want to take home (willingness-to-take-home) averaged across participants for each of the three sessions and of two emotional states. These behavioral measures showed a similar pattern of effects as the subjective ratings. There was no main effect of emotional state on sip size ($F(1, 66) = .26$, $p = .613$, $\eta_p^2 = .004$) and willingness-to-take-home ($F(1, 69) = .35$, $p = .557$, $\eta_p^2 = .005$), but significant interactions between emotional state and soups on both sip size ($F(1, 198) = 7.59$, $p = .006$, $\eta_p^2 = .126$) and willingness-to-take-home ($F(1, 345) = 11.16$, $p = .001$, $\eta_p^2 = .036$). Similar to valence and EsSense25 ratings, sip size and willingness-to-take-home were lower for novel soup than for familiar soup in the negative emotional state, while there was no difference between soups in the positive emotional state. This was corroborated by post-hoc comparisons. No significant three-way interactions were found for both measures, indicating a stable effect of emotion on familiarity across sessions. Table 5-1 shows that also the main effect of soup, and the lack of effect of state*session that were observed for valence ratings, and most of the Essense25 ratings are mirrored in the patterns of sip size and willingness-to-take-home *(hypothesis 5)*.

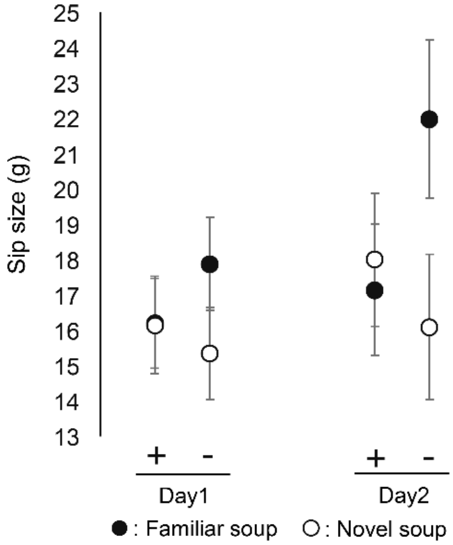


Figure 5-5. Mean sip size of familiar and novel soup by positive emotional group (+) and negative emotional group (-) in Day 1 and Day 2–2. Error bars indicate standard error of the mean.

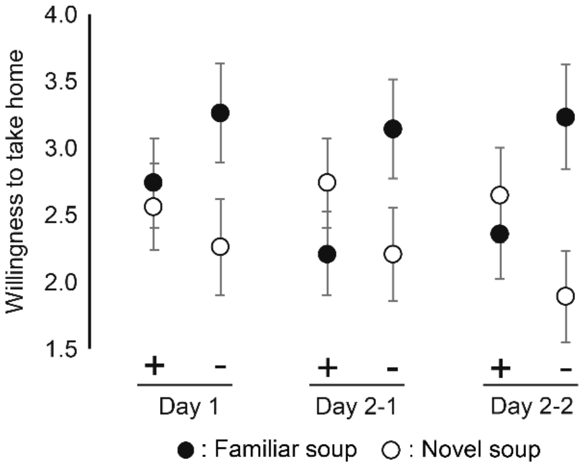


Figure 5-6. Mean willingness-to-take-home of familiar and novel soup by positive emotional group (+) and negative emotional group (-) in Day 1, Day 2–1, and Day 2–2. Error bars indicate standard error of the mean.

Table 5-1. Summary of the statistical data obtained with a mixed model analysis for each of dependent variable.

Dependent Variables	Soup (familiar, novel)	State (positive, negative)	Session (Day 1, Day 2-1, Day 2-2)	Soup x state	State x session	Soup x session	Soup x state x session
Valence	$F(1, 345) = 27.80$, $p < .001$, $\eta_p^2 = .112$	$F(1, 69) = .09$, $p = .764$, $\eta_p^2 = .001$	$F(2, 345) = 4.90$, $p = .080$, $\eta_p^2 = .122$	$F(1, 345) = 20.90$, $p < .001$, $\eta_p^2 = .087$	$F(2, 345) = .11$, $p = .896$, $\eta_p^2 = .003$	$F(2, 345) = 1.12$, $p = .328$, $\eta_p^2 = .039$	$F(2, 345) = .45$, $p = .635$, $\eta_p^2 = .016$
Sip size[1]	$F(1, 198) = 5.11$, $p = .025$, $\eta_p^2 = .089$	$F(1, 66) = .26$, $p = .613$, $\eta_p^2 = .004$	$F(1, 198) = 5.18$, $p = .024$, $\eta_p^2 = .045$	$F(1, 198) = 7.59$, $p = .006$, $\eta_p^2 = .126$	$F(1, 198) = .37$, $p = .541$, $\eta_p^2 = .003$	$F(1, 198) = .52$, $p = .471$, $\eta_p^2 = .014$	$F(1, 198) = 1.59$, $p = .209$, $\eta_p^2 = .043$
Willingness-to-take-home	$F(1, 345) = 5.02$, $p = .026$, $\eta_p^2 = .017$	$F(1, 69) = .35$, $p = .557$, $\eta_p^2 = .005$	$F(2, 345) = .28$, $p = .753$, $\eta_p^2 = .040$	$F(1, 345) = 11.16$, $p = .001$, $\eta_p^2 = .036$	$F(2, 345) = .05$, $p = .955$, $\eta_p^2 = .007$	$F(2, 345) = .36$, $p = .697$, $\eta_p^2 = .021$	$F(2, 345) = .40$, $p = .670$, $\eta_p^2 = .023$
Active	$F(1, 345) = 3.59$, $p = .059$, $\eta_p^2 = .027$	$F(1, 69) = .52$, $p = .474$, $\eta_p^2 = .007$	$F(2, 345) = 1.93$, $p = .146$, $\eta_p^2 = .027$	$F(1, 345) = 2.01$, $p = .157$, $\eta_p^2 = .015$	$F(2, 345) = .31$, $p = .733$, $\eta_p^2 = .005$	$F(2, 345) = 1.53$, $p = .586$, $\eta_p^2 = .038$	$F(2, 345) = .43$, $p = .651$, $\eta_p^2 = .011$
Adventurous	$F(1, 345) = 86.37$, $p < .001$, $\eta_p^2 = .336$	$F(1, 69) = .14$, $p = .705$, $\eta_p^2 = .002$	$F(2, 345) = .35$, $p = .080$, $\eta_p^2 = .008$	$F(1, 345) = .00$, $p = .983$, $\eta_p^2 < .001$	$F(2, 345) = 1.64$, $p = .195$, $\eta_p^2 = .036$	$F(2, 345) = .54$, $p = .219$, $\eta_p^2 = .012$	$F(2, 345) = .30$, $p = .739$, $\eta_p^2 = .007$
Aggressive	$F(1, 345) = 58.07$, $p < .001$, $\eta_p^2 = .248$	$F(1, 69) = 3.41$, $p = .069$, $\eta_p^2 = .047$	$F(2, 345) = 1.73$, $p = .179$, $\eta_p^2 = .040$	$F(1, 345) = 2.09$, $p = .149$, $\eta_p^2 = .012$	$F(2, 345) = .14$, $p = .873$, $\eta_p^2 = .003$	$F(2, 345) = 1.46$, $p = .234$, $\eta_p^2 = .033$	$F(2, 345) = .19$, $p = .829$, $\eta_p^2 = .004$
Bored	$F(1, 345) = 5.95$, $p = .015$, $\eta_p^2 = .043$	$F(1, 69) = .54$, $p = .464$, $\eta_p^2 = .008$	$F(2, 345) = .01$, $p = .990$, $\eta_p^2 < .001$	$F(1, 345) = 4.54$, $p = .034$, $\eta_p^2 = .043$	$F(2, 345) = .45$, $p = .639$, $\eta_p^2 = .007$	$F(2, 345) = .37$, $p = .693$, $\eta_p^2 = .009$	$F(2, 345) = 1.36$, $p = .259$, $\eta_p^2 = .031$
Clam	$F(1, 345) = 45.03$, $p < .001$, $\eta_p^2 = .244$	$F(1, 69) = .02$, $p = .889$, $\eta_p^2 < .001$	$F(2, 345) = 1.37$, $p = .257$, $\eta_p^2 = .022$	$F(1, 345) = .27$, $p = .601$, $\eta_p^2 = .002$	$F(2, 345) = 1.82$, $p = .163$, $\eta_p^2 = .029$	$F(2, 345) = 1.62$, $p = .199$, $\eta_p^2 = .037$	$F(2, 345) = .21$, $p = .812$, $\eta_p^2 = .005$
Disgust	$F(1, 345) = 34.68$, $p < .001$, $\eta_p^2 = .127$	$F(1, 69) = 1.80$, $p = .184$, $\eta_p^2 = .025$	$F(2, 345) = 1.60$, $p = .203$, $\eta_p^2 = .058$	$F(1, 345) = 7.56$, $p = .006$, $\eta_p^2 = .031$	$F(2, 345) = .13$, $p = .878$, $\eta_p^2 = .005$	$F(2, 345) = .16$, $p = .853$, $\eta_p^2 = .006$	$F(2, 345) = .05$, $p = .953$, $\eta_p^2 = .002$
Enthusiastic	$F(1, 345) = .02$, $p = .885$, $\eta_p^2 < .001$	$F(1, 69) = .00$, $p = .998$, $\eta_p^2 < .001$	$F(2, 345) = 1.12$, $p = .327$, $\eta_p^2 = .023$	$F(1, 345) = 4.68$, $p = .031$, $\eta_p^2 = .023$	$F(2, 345) = .67$, $p = .512$, $\eta_p^2 = .014$	$F(2, 345) = .83$, $p = .439$, $\eta_p^2 = .029$	$F(2, 345) = .14$, $p = .870$, $\eta_p^2 = .005$
Free	$F(1, 345) = 3.44$, $p = .064$, $\eta_p^2 = .034$	$F(1, 69) = .25$, $p = .616$, $\eta_p^2 = .004$	$F(2, 345) = .02$, $p = .985$, $\eta_p^2 < .001$	$F(1, 345) = 3.44$, $p = .064$, $\eta_p^2 = .034$	$F(2, 345) = .55$, $p = .580$, $\eta_p^2 = .008$	$F(2, 345) = .46$, $p = .632$, $\eta_p^2 = .008$	$F(2, 345) = 1.17$, $p = .313$, $\eta_p^2 = .021$
Good	$F(1, 345) = 16.36$, $p < .001$, $\eta_p^2 = .099$	$F(1, 69) = .14$, $p = .711$, $\eta_p^2 = .002$	$F(2, 345) = 2.37$, $p = .095$, $\eta_p^2 = .036$	$F(1, 345) = 10.10$, $p = .002$, $\eta_p^2 = .064$	$F(2, 345) = .16$, $p = .854$, $\eta_p^2 = .002$	$F(2, 345) = .11$, $p = .894$, $\eta_p^2 = .003$	$F(2, 345) = .16$, $p = .855$, $\eta_p^2 = .005$

---

[1] There are only two levels of session for sip size (Day 1 and Day 2-2)

Good natured	F (1, 345) = 23.99, p < .001, η_p^2 = .143	F(1, 69) = .90, p = .345, η_p^2 = .013	F(2, 345) = 5.74, p = .004, η_p^2 = .091	F(1, 345) = 1.23, p = .267, η_p^2 = .009	F(2, 345) = 1.11, p = .331, η_p^2 = .019	F(2, 345) = .41, p = .667, η_p^2 = .009	F(2, 345) = .92, p = .401, η_p^2 = .021
Guilty	F (1, 345) = 13.14, p < .001, η_p^2 = .094	F(1, 69) = 2.41, p = .125, η_p^2 = .034	F(2, 345) = 4.10, p = .017, η_p^2 = .057	F(1, 345) = 5.82, p = .016, η_p^2 = .044	F(2, 345) = .21, p = .808, η_p^2 = .003	F(2, 345) = 1.90, p = .151, η_p^2 = .044	F(2, 345) = .14, p = .871, η_p^2 = .003
Happy	F (1, 345) = 24.39, p < .001, η_p^2 = .142	F(1, 69) = .23, p = .631, η_p^2 = .003	F(2, 345) = .93, p = .397, η_p^2 = .015	F(1, 345) = 18.08, p < .001, η_p^2 = .109	F(2, 345) = .48, p = .621, η_p^2 = .008	F(2, 345) = .33, p = .722, η_p^2 = .008	F(2, 345) = .19, p = .828, η_p^2 = .005
Interested	F (1, 345) = .39, p = .532, η_p^2 = .002	F(1, 69) = .20, p = .653, η_p^2 = .003	F(2, 345) = 2.34, p = .098, η_p^2 = .038	F(1, 345) = 3.81, p = .052, η_p^2 = .023	F(2, 345) = .64, p = .527, η_p^2 = .011	F(2, 345) = .02, p = .985, η_p^2 < .001	F(2, 345) = .51, p = .601, η_p^2 = .016
Joyful	F (1, 345) = 5.79, p = .017, η_p^2 = .034	F(1, 69) = .62, p = .433, η_p^2 = .009	F(2, 345) = 2.14, p = .120, η_p^2 = .044	F(1, 345) = 4.61, p = .033, η_p^2 = .027	F(2, 345) = .95, p = .387, η_p^2 = .020	F(2, 345) = .55, p = .577, η_p^2 = .012	F(2, 345) = .11, p = .899, η_p^2 = .002
Loving	F (1, 345) = 23.52, p < .001, η_p^2 = .180	F(1, 69) = .00, p = .973, η_p^2 = .	F(2, 345) = .17, p = .840, η_p^2 = .003	F(1, 345) = 4.38, p = .037, η_p^2 = .039	F(2, 345) = 3.50, p = .031, η_p^2 = .057	F(2, 345) = .39, p = .680, η_p^2 = .006	F(2, 345) = .94, p = .393, η_p^2 = .015
Mild	F (1, 345) = 52.00, p < .001, η_p^2 = .258	F(1, 69) = .19, p = .668, η_p^2 = .003	F(2, 345) = .45, p = .639, η_p^2 = .008	F(1, 345) = 2.53, p = .113, η_p^2 = .017	F(2, 345) = .30, p = .745, η_p^2 = .005	F(2, 345) = 2.17, p = .116, η_p^2 = .052	F(2, 345) = .43, p = .651, η_p^2 = .011
Nostalgic	F (1, 345) = 109.62, p < .001, η_p^2 = .398	F(1, 69) = .94, p = .335, η_p^2 = .013	F(2, 345) = .59, p = .558, η_p^2 = .013	F(1, 345) = 3.40, p = .066, η_p^2 = .020	F(2, 345) = 1.84, p = .160, η_p^2 = .039	F(2, 345) = 2.34, p = .098, η_p^2 = .050	F(2, 345) = .45, p = .635, η_p^2 = .010
Pleasant	F (1, 345) = 9.56, p = .002, η_p^2 = .049	F(1, 69) = .00, p = .984, η_p^2 < .001	F(2, 345) = .26, p = .774, η_p^2 = .006	F(1, 345) = 11.75, p = .001, η_p^2 = .060	F(2, 345) = .09, p = .917, η_p^2 = .002	F(2, 345) = .11, p = .901, η_p^2 = .003	F(2, 345) = .25, p = .778, η_p^2 = .007
Satisfied	F (1, 345) = 24.79, p < .001, η_p^2 = .116	F(1, 69) = .00, p = .974, η_p^2 < .001	F(2, 345) = 3.49, p = .032, η_p^2 = .067	F(1, 345) = 1.42, p = .234, η_p^2 = .007	F(2, 345) = .81, p = .445, η_p^2 = .016	F(2, 345) = .48, p = .617, η_p^2 = .016	F(2, 345) = .97, p = .380, η_p^2 = .032
Secure	F (1, 345) = 33.69, p < .001, η_p^2 = .194	F(1, 69) = .51, p = .478, η_p^2 = .007	F(2, 345) = 1.53, p = .217, η_p^2 = .029	F(1, 345) = 1.32, p = .252, η_p^2 = .009	F(2, 345) = 1.89, p = .153, η_p^2 = .036	F(2, 345) = 1.12, p = .327, η_p^2 = .021	F(2, 345) = 2.31, p = .101, η_p^2 = .043
Tame	F (1, 345) = 32.04, p < .001, η_p^2 = .212	F(1, 69) = 1.05, p = .309, η_p^2 = .015	F(2, 345) = .26, p = .769, η_p^2 = .004	F(1, 345) = .09, p = .762, η_p^2 = .001	F(2, 345) = .38, p = .682, η_p^2 = .006	F(2, 345) = .35, p = .704, η_p^2 = .008	F(2, 345) = 2.31, p = .101, η_p^2 = .049
Understanding	F (1, 345) = 20.71, p < .001, η_p^2 = .190	F(1, 69) = .01, p = .913, η_p^2 < .001	F(2, 345) = 5.37, p = .005, η_p^2 = .067	F(1, 345) = 3.53, p = .061, η_p^2 = .038	F(2, 345) = 2.97, p = .052, η_p^2 = .038	F(2, 345) = .05, p = .950, η_p^2 = .001	F(2, 345) = 1.68, p = .188, η_p^2 = .031
Warm	F (1, 345) = 22.10, p < .001, η_p^2 = .155	F(1, 69) = .35, p = .554, η_p^2 = .005	F(2, 345) = 1.76, p = .174, η_p^2 = .024	F(1, 345) = 5.42, p = .020, η_p^2 = .043	F(2, 345) = 2.19, p = .113, η_p^2 = .029	F(2, 345) = .67, p = .513, η_p^2 = .017	F(2, 345) = 1.03, p = .358, η_p^2 = .025

Wild	*F* (1, 345) = 80.82, *p* < .001, η_p^2 = .316	*F*(1, 69) = 1.23, *p* = .272, η_p^2 = .017	*F*(2, 345) = 2.49, *p* = .084, η_p^2 = .056	*F*(1, 345) = .52, *p* = .471, η_p^2 = .003	*F*(2, 345) = .31, *p* = .732, η_p^2 = .007	*F*(2, 345) = .36, *p* = .697, η_p^2 = .008	*F*(2, 345) = .03, *p* = .974, η_p^2 = .001
Worried	*F* (1, 345) = 10.32, *p* = .001, η_p^2 = .077	*F*(1, 69) = 2.48, *p* = .120, η_p^2 = .035	*F*(2, 345) = 1.37, *p* = .257, η_p^2 = .021	*F*(1, 345) = 2.57, *p* = .110, η_p^2 = .020	*F*(2, 345) = 4.44, *p* = .012, η_p^2 = .066	*F*(2, 345) = .57, *p* = .568, η_p^2 = .012	*F*(2, 345) = .38, *p* = .682, η_p^2 = .008

The significant main effects and interactions were highlighted in light grey. Variables below the dashed lines reflect the EsSence25 variables. [a]There are only two levels of session for sip size (Day 1 and Day 2–2).

Table 5-2. Summary of LSD post-hoc comparisons that elucidate significant soup x state interactions.

Soup vs. state	Familiar Positive vs. Novel Positive	Familiar Negative vs. Novel Negative	Familiar Positive vs. Familiar Negative	Novel Positive vs. Novel Negative
Valence	*p* = .706	*p* < .001	*p* = .005	*p* = .035
Sip size	*p* = .946	*p* = .014	*p* = .060	*p* = .517
Willingness-to-take-home	*p* = .449	*p* < .001	*p* = .006	*p* = .060
Good	*p* = .650	*p* < .001	*p* = .033	*p* = .226
Happy	*p* = .738	*p* < .001	*p* = .147	*p* = .007
Pleasant	*p* = .855	*p* < .001	*p* = .057	*p* = .064

The significant effects are highlighted in light grey. Variables below the dashed lines reflect the EsSence25 variables.

Discussion

The present study investigated the effect of emotional state (positive and negative) on valence and EsSense25 ratings, reported willingness-to-take-home, and sip size for novel and familiar soups, both at the time of emotion induction (Day 1) as well as at two recording sessions a week after (Day 2-1, without tasting, and Day 2-2, with tasting).

At Day 1, participants tasted and rated the soups for the first time, after an either positive or negative emotion induction procedure. For Day 1, we expected that overall experienced food pleasantness, as reflected in the valence ratings and the EsSense25, would be lower when tasting soups in a negative emotional state than in a positive emotional state, and that this effect would be stronger for the novel soup than for the familiar soup. We indeed observed that participants with a negative emotional state rated lower valence for the novel soup than for the familiar soup, whereas there was no difference between soups in the positive condition. However, the lack of a main effect of emotional state indicated that this was not merely due to a general lower valence in the negative condition. Rather, familiar soup was rated more positively in the negative emotional condition than in the positive emotional condition. For EsSense25, this pattern was found for three positive emotions (*Happy*, *Pleasant* and *Good*) and an additional three when a more liberal criterion of significance was taken (*Warm*, *Joyful*, and *Enthusiastic*) and three negative emotional terms (only without Bonferoni correction: *Bored*, *Disgusted*, and *Guilty*). These results force us to reject hypothesis 1 – we did not find that negative emotional state decreased experienced food pleasantness in general, but partly supported hypothesis 2 – we found that negative emotional state decreased food pleasantness particularly for novel foods, where familiar food, contrary to our expectation, rather increased in food pleasantness. Our results are consistent with a comforting effect of a familiar taste in a stressful situation.

In the negative emotional group, we found that familiar soup was preferred over novel soup. In general, familiar foods are reported to be preferred over unfamiliar food. Fenko et al. (2015b) investigated participants' hedonic responses to various familiar and unfamiliar soy product images and found higher liking scores for familiar products, as well as a more positive expectation of the familiar products' taste. Consistent with this, Toet et al. (2019c) found that Asian and Western participants rated food from their own culture as more positive. Our study shows that this tendency may be especially strong in stressful situations. This is also suggested by a study from Locher et al. (2005). They asked participants to bring foods that "made them feel good" or "provided them comfort" and to explain why this was so. They concluded that people consume familiar foods to relieve feelings of distress and anxiety, and that novel foods cannot fulfill this need because they tend to evoke more feelings of anxiety. Other studies report that individuals in depressed moods show a preference for and consume palatable well-known "comfort foods" to alleviate their negative feelings (Macht, 2008; Singh, 2014). Also, it is reported in a review that familiar foods represent the sense of perceived "comfort" while it is absent with novel foods because of a lack of knowledge of them (Aldridge et al., 2009).

We hypothesized that the differential effects of emotional state on novel and familiar soups would be stronger a week later when the actual taste of the soup is not available (hypothesis 3). This pattern could be observed in most measures. For valence ratings, the difference between familiar and novel soup in Day 2-1 tended to be larger for participants that had been under negative stress than for participants from the positive emotional condition. The EsSense25 measures showed similar effects for certain positive emotions, such as "*Good*", "*Joyful*", "*Happy*", "*Warm*", "*Pleasant*", and "*Enthusiastic*", as well as sip size. However, significant three-way interactions between soup, state and session were far from significant for any of the measures. Thus, we conclude that the differential effect of emotion on experiencing familiar and novel soup on Day 1, when the emotions were induced, remained the same a week later, therewith rejecting hypothesis 3.

We expected that when participants would taste the soups again (Day2-2), this would reduce the effects of memory (hypothesis 4). However, the lack of three-way interactions between soup, state and session showed this not to be the case - even after tasting the soups again, the interactive effect of soup and emotional state remained the same for all measures. Hypothesis 4 was therefore rejected. Our data showed that the interactive effect of emotional state and familiarity (soup) is robust. For ratings, this may have been caused by participants being inclined to give similar answers as they did before. De Wijk et al. (2019) pointed out that memories of previous encounters with the same test food may induce the use of similar ratings in new encounters. However, the finding that our implicit measure of sip size produced the same results argues against this explanation in our study.

We used different measures to evaluate food experience from various angles. Valence rating, sip size and willingness-to-take home showed similar patterns of effects of soup, state and session. This pattern was also seen in emotions probed in the EsSense25. Our hypothesis 5 was thus confirmed.

Overall, we found that the results related to participants' food evaluations (valence, EsSense25, willingness-to-take-home, and sip size) did not completely follow our hypotheses. In fact, no main effects were observed of emotional state, though we did find significant interactions between emotional state and food novelty in all measures. The results showed that in the negative emotional condition, familiar foods were rated more positively than novel foods, whereas they were rated the same in a positive emotional state. We had expected familiar foods to be more stable than novel foods across emotional conditions because of a long-term emotional association in memory in the participants. The fact that a food is familiar, can be taken to mean that it is 'safe', and thus in general more positive than novel foods. In the negative, stressed condition, individuals may have been more sensitive to any potential threats, resulting in an increased avoidance (or negative valence). On the other hand, when in a positive state, there is no reason to activate the threat awareness or avoidance mechanism, and individuals do not avoid food just because it is unfamiliar. The ratings in the positive emotional group may be mainly based on the smell and taste and not so much affected by the fact that they have not experienced it before. For future studies, it would be of interest to investigate if, and how food neophobia affects these interactions between emotional state and food novelty on experienced food pleasantness.

The results in this study implicate that one should introduce a new product (or a novel food) in a situation where people are not stressed. Introducing a new product to consumers who are likely to be stressed (e.g. in a hospital) would not be recommended as it may affect food pleasantness negatively and for a long period. Once people have experienced a new product in a stressed state, the negative effect is robust at least for one week. Therefore, a positive recommendation would be to let people taste a new product when they are in a positive mood, e.g. at a festival or after a happy movie.

Conclusion

This study evaluated the interaction effects between emotional states and food novelty on food experiences in terms of valence and EsSense25 ratings, willingness-to-take-home and sip size, both during initial tasting of two soups and a week later. We showed that the emotional state affected all these measures in a similar way, with low experienced pleasantness for novel food and high pleasantness for familiar food in the negative compared to the positive emotional group, in which no differences in pleasantness were found. Also, the effects of emotional state proved to be robust over time (one week later in this study). Our findings in this study provided relevant insights for food industries and restaurants for introducing their new products to consumers and for hospitals and care institutions for providing medication or food supplements.

Acknowledgements

The authors thank Esther Bosch, Wouter Korteling, and Benjamin de Graaff for their assistance with conducting the experiment and analyzing data.

Chapter 6

CROCUFID: A Cross-Cultural Food Image Database for Research on Food Elicited Affective Responses

This chapter is published as:
Toet, A., Kaneko, D., De Kruijf, I., Ushiama, S., Van Schaik, M.G., Brouwer, A.M., Kallen, V., and Van Erp, J.B.F. (2019). CROCUFID: A cross-cultural food image database for research on food elicited affective responses. Frontiers in psychology 10, 58.

Abstract

We present CROCUFID: a CROss-CUltural Food Image Database that currently contains 840 images, including 479 food images with detailed metadata and 165 images of non-food items. The database includes images of sweet, savory, natural, and processed food from Western and Asian cuisines. To create sufficient variability in valence and arousal we included images of food with different degrees of appetitiveness (fresh, unfamiliar, molded or rotten, spoiled, partly consumed). We used a standardized photographing protocol, resulting in high resolution images depicting all food items on a standard background (a white plate), seen from a fixed viewing (45 deg) angle. CROCUFID is freely available under the CC-By Attribution 4.0 International license and hosted on the OSF repository. The advantages of the CROCUFID database over other databases are its (1) free availability, (2) full coverage of the valence - arousal space, (3) use of standardized recording methods, (4) inclusion of multiple cuisines and unfamiliar foods, (5) availability of normative and demographic data, (6) high image quality and (7) capability to support future (e.g. virtual and augmented reality) applications. Individuals from the UK (N=266), North-America (N=275) and Japan (N=264) provided normative ratings of valence, arousal, perceived healthiness, and desire-to-eat using visual analogue scales (VAS). In addition, for each image we computed 17 characteristics that are known to influence affective observer responses (e.g., texture, regularity, complexity, colorfulness). Significant differences between groups and significant correlations between image characteristics and normative ratings were in accordance with previous research, indicating the validity of CROCUFID. We expect that CROCUFID will facilitate comparability across studies and advance experimental research on the determinants of food-elicited emotions. We plan to extend CROCUFID in the future with images of food from a wide range of different cuisines and with non-food images (for applications in for instance neuro-physiological studies). We invite researchers from all parts of the world to contribute to this effort by creating similar image sets that can be linked to this collection, so that CROCUFID will grow into a truly multicultural food database.

Introduction

Visual cues constitute a primary sensory input that allows predictions about the edibility and palatability of food. Through learning, visual food characteristics can become secondary reinforcers that affect eating-related behavior. The human brain has specific regions involved in the appetitive and affective processing of visual presentations of food stimuli (Killgore et al., 2003). The sight of food elicits a wide range of physiological, emotional and cognitive responses (van der Laan et al., 2011). Previous research has shown that viewing pictures of food not only activates the visual cortex, but also brain areas that code how food actually tastes (the insula/operculum) and the reward values of tasting it (the orbitofrontal cortex; (Simmons et al., 2005). Food images are therefore typically considered as useful proxies for the real thing (e.g., (Foroni et al., 2013; Blechert et al., 2014b; Miccoli et al., 2014; Miccoli et al., 2016)). Food pictures are increasingly used as stimuli in research on the factors underlying affective and appetitive responses to foods (e.g., in online experiments (Blechert et al., 2014b; Jensen et al., 2016)), especially when the experimental paradigm limits actual consumption (e.g., experiments involving physiological measures such as EEG and fMRI; (Killgore et al., 2003; Piqueras-Fiszman et al., 2014)).

The sensory characteristics and evoked emotions of food are crucial factors in predicting a consumer’s food preference and therefore in developing new products (Dalenberg et al., 2014; Gutjar et al., 2015b). Hedonic ratings alone do not predict food choice behavior accurately (Zandstra and El-Deredy, 2011; Griffioen-Roose et al., 2013b). Emotions have an incremental predictive value over hedonic ratings alone, so that the best prediction can be obtained by combining both measures (Dalenberg et al., 2014; Samant and Seo, 2018). The relationship between food and emotions appears to be bidirectional: emotions influence eating behavior, while eating behavior also affects the consumer’s emotional state (Desmet and Schifferstein, 2008). Assessing emotional responses to foods could reveal product attributes that are a valuable source of information for product development and marketing going beyond traditional sensory and acceptability measurements (Thomson et al., 2010). Given the high failure rate of new food products in the market (between 60 and 80%: (Köster and Mojet, 2007)) and the fact that there is often only little difference in quality, price and design of different food products, knowledge about the affective responses of food products appears to be important for the food industry (Köster, 2003; Schifferstein et al., 2013). Therefore, it is important to obtain valid and reliable measurements of food-evoked emotions. Pictures are easy to present and have been widely used to investigate the different factors underlying affective responses in general (Lang et al., 1993) and appetitive responses to foods (Blechert et al., 2014a; Hebert et al., 2015). To facilitate multi-disciplinary and cross-cultural studies on food-related emotions we present a publicly available database containing images of a wide range of various foods from different cuisines, together with normative observer ratings.

Despite their recognized importance for research, only a few food image databases have been made publicly available to support research on human eating behavior (Chen et al., 2009; Foroni et al., 2013; Blechert et al., 2014b; Miccoli et al., 2014; Charbonnier et al., 2016) and they fall short on one or more of the following important criteria: (1) the public availability and ease of reproducibility, (2) the coverage of the full valence and arousal space, (3) the use of standardized recording methods, (4) the use of multiple cuisines and unfamiliar foods, (5) the availability of normative and demographic data (6) the image quality and capability for future applications. We will discuss the importance of these criteria in the following sections.

Public availability and ease of reproducibility

Researchers often select food images based on their personal preferences. As a result, the image sets are typically highly variable and difficult to compare, re-use or replicate (Pashler and Wagenmakers, 2012; Charbonnier et al., 2016). While some studies used their own set of images (Manzocco et al., 2013; Piqueras-Fiszman et al., 2014; Charbonnier et al., 2015), others collected stimuli from the internet (Foroni et al., 2013; Blechert et al., 2014b; Miccoli et al., 2014; Hebert et al., 2015). Also, most studies report little information about the used pictures and do not make the used image sets available to other researchers. It has been recognized that the comparability of research findings across studies using food pictures would benefit from the availability of public databases with standardized images (Foroni et al., 2013; Blechert et al., 2014b; Charbonnier et al., 2016).

Coverage of the affective space

Existing food image databases typically focus on appetitive and positively valenced food pictures and lack aversive or negatively valenced ones. As a result, they cover only a limited part of the valence-arousal space, which significantly reduces their value for studies on food-evoked emotions. This is a result of the fact that food-related research has typically investigated positive and familiar food products while neglecting negative and unfamiliar ones, thus ignoring emotions and behaviors like disgust and rejection. This may suffice in product development and consumer research, but not in research on emotional responses to less liked and disliked food (Piqueras-Fiszman et al., 2014). Although positive food related emotional states are far more common, it is unavoidable that consumers sometimes experience negative food-related emotions (Desmet and Schifferstein, 2008; King et al., 2010; Cardello et al., 2012; Manzocco et al., 2013; Ng et al., 2013; Spinelli et al., 2015). Image databases intended to support research on human food-related behavior and to understand the full variety of emotional responses to food should therefore also include images of negatively-valenced and unfamiliar products.

Standardized images

Standardization across image characteristics (e.g., background, color, brightness, contrast, size etc.) will enhance the reliability and validity of experimental results. However, most of the current databases contain unstandardized images of food items, sometimes simply collected from the internet and pasted on white backgrounds. The resulting large variation in image characteristics (e.g., contrast, brightness), the different angles at which foods are depicted, the variation in magnifications, the lack of a visual reference such as a plate together with the lack of shadows may compromise the reliability and validity of the experimental results from studies in which these images are used (Charbonnier et al., 2016).

Multiple cuisines and unfamiliar foods

Cultural background and familiarity are other factors that strongly influence food-related emotional responses (van Zyl and Meiselman, 2015) and determine food choice (Rozin, 1996). Databases should preferably include images representing a wide range of different food types (natural, processed) from a variety of cuisines. Existing food image databases include only a small range of (typically Western) cuisines, which limits their value for cross-cultural studies. Hence, there is a need for a food image database that includes food from different cuisines and has been validated across different nations, such as Latin-American, African, and Middle-Eastern.

Normative and demographic data

The inclusion of different food types in a food image database enables the selection of food with different colors, caloric content, macronutrients, readiness to eat, flavor, nutritional composition, healthiness, and familiarity and the distinction of different classes such as vegetables, meat-containing dishes, fruits, and snacks (e.g. (Blechert et al., 2014b)).

A validated food-image database should preferably also include data on individual differences that are known to influence normative subjective ratings like age (Hoogeveen et al., 2015), gender (e.g., (Cepeda-Benito et al., 2003; Toepel et al., 2012)), body mass index (BMI; e.g. (Berthoud and Zheng, 2012; Toepel et al., 2012)), food preferences or diets (e.g., (Stockburger et al., 2009)) and psychophysiological state (e.g., (Hoefling et al., 2009)). Also, the measures included in the validation of the database should make a careful distinction between liking (the hedonic appraisal of food) and wanting (the motivation to consume: (Berridge, 2004)). In addition, the inclusion of images of artificial non-food objects (either related to food - like tableware or cutlery - or unrelated to food - like office supplies or toys), and organic non-food objects (like plants and flowers) is recommended to support brain studies (van der Laan et al., 2011).

Image quality and suitability for different applications

Item-only(transparent) food images can be used in different experimental techniques like simulations of different ambient environments and multisensory cues. Recent studies have shown that the sensorial and hedonic experience of food is significantly influenced by the color (Genschow et al., 2012; Piqueras-Fiszman et al., 2012; Piqueras-Fiszman et al., 2013; Stewart and Goss, 2013; Michel et al., 2015b; Tu et al., 2016; Chen et al., 2018a), shape (Piqueras-Fiszman et al., 2012; Stewart and Goss, 2013; Chen et al., 2018a) and size (Van Ittersum and Wansink, 2012; Wansink and van Ittersum, 2013) of the plateware that is used for serving. In real-world settings, contextual factors such as ambience (Stroebele and De Castro, 2004), room color (Meléndez-Martínez et al., 2005; Oberfeld et al., 2009; Spence et al., 2014; Schifferstein et al., 2017), background sounds and music (McCarron and Tierney, 1989; Spence and Shankar, 2010; Crisinel et al., 2012; Fiegel et al., 2014), ambient temperature (Herman, 1993), the color and intensity of ambient lighting (Oberfeld et al., 2009; Suk et al., 2012; Hasenbeck et al., 2014; Spence et al., 2014; Cho et al., 2015) and ambient scents (Spence, 2015a) are also known to modulate the assessment and consumption of food and drinks (for reviews see: (Spence and Piqueras-Fiszman, 2014; Spence, 2018)). It has been suggested that these effects reflect a transfer of sensations through cross-modal correspondences (Spence, 2011a). However, the exact nature of these effects and the way in which they interact is still a topic of research. Acquiring further knowledge about the way in which multisensory contextual and ambient cues interact and affect human food related behavior will be of great value in retail and restaurant settings and may help to improve food experience and consumption behavior (e.g., to fight obesity, to enhance the consumption experience of elderly or people in hospital and care facilities, etc.). Virtual Reality (VR: (Gorini et al., 2010b; Nordbo et al., 2015b; Ung et al., 2018a)), Augmented Reality (AR: (Narumi et al., 2012b; Pallavicini et al., 2016)), Mixed Reality (MR: (Huisman et al., 2016)) or other immersive technologies (Bangcuyo et al., 2015; Liu et al., 2018) appear to be promising tools for this kind of research (see also: (Spence and Piqueras-Fiszman, 2014)). The extension of VR and AR systems with novel multisensory (taste, smell, tactile) interfaces can further enhance the perceived reality of food imagery (Braun et al., 2016) and will provide researchers control over the various inputs that determine a given food experience (Velasco et al., 2018b). This may significantly increase the effectiveness of these systems for studies on food-related emotions and behavior, personal health and wellbeing (Comber et al., 2014b; Obrist et al., 2016). Multisensory HCI systems may for instance be used to match the visual, auditory and olfactory characteristics of a simulated table or restaurant setting to food (e.g., (Reinoso Carvalho et al., 2016)). Transparent item-only images can be overlaid on images of plates with different shapes, colors and textures to study the effects of the visual characteristics of the plate on the appraisal of food. To study the effects of the visual characteristics of the environment, these plated food images can in turn be overlaid on images, movies or VR renderings of different backgrounds (e.g., images showing plated food placed on tables covered with different tablecloths or on different natural surfaces, movies showing environments with different lighting characteristics, etc.). To study the effects of ambient sound, (dynamic) lighting or social presence, the plated food images can be overlaid on movies showing dynamic environments with different ambient characteristics. Adding smells, tastes or tactile stimulation to the image presentation may serve to further enhance the realism of this type of studies (Velasco et al., 2018b). Using item-only images in augmented reality settings will provide an efficient way to study the effects of ambient characteristics on human response to many different types of food (Nishizawa et al., 2016).

The CROCUFID approach

To complement the currently available food image databases and to further support systematic neuroscientific and behavioral studies on the emotional impact of food we present CROCUFID: a CROss-CUltural Food Image Database with high-resolution images of different types of food from various (currently mainly Western and Asian) cuisines, together with normative ratings (by participants from the United Kingdom, North-America and Japan) of valence, arousal, desire-to-eat, perceived healthiness, and familiarity, complemented with computational measures of physical food properties that are known (or a priori considered likely) to influence food experience. To make it useful for brain (e.g., fMRI) studies, the dataset also contains images of non-food objects. To cover the full valence-arousal space, CROCUFID includes images of food with different degrees of appetitiveness (fresh, unfamiliar, molded or rotten, contaminated, partly consumed). The inclusion of food types from different (Western and Asian) cuisines allows the images to be used in a culturally diverse population. The images resemble the viewing of a plate of food on a table during meal time and will therefore be a useful tool to assess emotional responses in studies that limit actual consumption (e.g., experiments involving physiological measures such as EEG and fMRI). To afford the use of CROCUFID in human-computer interaction (HCI) studies, all images in the dataset are also provided as item-only (transparent PNG) images. CROCUFID complements the F4H image collection since both were registered using the same photographing protocol (Charbonnier et al., 2016).

Related works

In this section we first describe the characteristics of some currently and publicly available food image databases that have been designed to support neuroscientific and behavioral research on human eating behavior and preferences. Then we review the databases that have been constructed to develop and train automatic food recognition and ingredient or recipe retrieval algorithms. To the best of our knowledge, these databases are currently the only ones publicly available that contain images of a wide range of different cuisines. However, they generally appear to be unsuitable for systematic research on human food-related behavior since they typically contain real-life images with largely varying backgrounds, taken from different points of view, with different scales and rotation angles and under varying lighting conditions. Next, we describe the value of CROCUFID for studies on the effects of environmental characteristics and background context on human food experience. Finally, we discuss how CROCUFID can be used to perform cross-cultural food studies.

Food image databases for human observer studies

Table 6-1 provides an overview of publicly available food image databases for human observer studies.

The Foodcast Research Image Database (FRIDa: (Foroni et al., 2013); see: https://foodcast.sissa.it/neuroscience) contains images of predominantly Western natural, transformed, and rotten food, natural and artificial non-food objects, animals, flowers and scenes, along with a description of several physical product properties (e.g., size, brightness and spatial frequency content) and normative ratings (by Italian participants) on several dimensions (including valence, arousal and familiarity). The items were collected from the internet, pasted on a white background and have a low resolution (530×530 pixels).

Table 6-1. Overview of food image databases for human observer studies.

Databases	Coverage of affective space	Recording methods	Cuisines	Availability of normative (and demographic) data	Remarks
FRIDa (Foroni et al. 2013)	Mainly positive valence	Not standardized	Mainly Western	Valence, Arousal, Familiarity (Italian)	• Low resolution (530×530 pixels) • Collected from Internet • Includes non-food images
Food-Pics (Blechert et al. 2014)	Mainly positive valence	Not standardized	Mainly Western	Valence, Arousal, Familiarity, Recognizability, Complexity, Palatability (German and North American)	• Low resolution (600×450 pixels) • Collected from Internet • No fixed background
OLAF (Miccoli et al. 2014; Miccoli et al. 2016)	Mainly positive valence	Not standardized	Mainly Western	Valence, Arousal, Dominance, Food Craving (Spanish)	• High resolution (4000×3000 pixels) • Includes some low valence images from IAPS • Includes non-food images
F4H (Charbonnier et al. 2016)	Mainly positive valence	Standardized	Mainly Western	Liking, Healthiness, Recognizability, Perceived Calories (Greek, Dutch, Scottish, German, Hungary, and Swedish)	• Resolution (3872×2592 pixels) • All images registered by the authors • Includes non-food images

The Food-pics database ((Blechert et al., 2014b); see: http://www.eat.sbg.ac.at/resources/food-pics) contains images of predominantly Western food types, together with normative ratings (by participants from German speaking countries and North America) on familiarity, recognizability, complexity, valence, arousal, palatability, and desire to eat. The items were collected from the internet, pasted on a white background and have a low resolution (600×450 pixels). Food-pics has been designed to support experimental research on food perception and eating behavior in general. The Open Library of Affective Foods (OLAF: (Miccoli et al., 2014; Miccoli et al., 2016); see: https://doi.org/10.5281/zenodo.10202) is a database of food pictures representing four different types of Western food (vegetables, fruit, sweet and salty high-fat foods), along with normative ratings (by Spanish students) of valence, arousal, dominance and food craving. The images have a high resolution (up to 4000×3000 pixels) and include food served in restaurants and homemade meals, and display non-food items in the background to increase their ecological value and to resemble the appearance of images from the International Affective Picture System (IAPS: (Lang et al., 2005)). The four selected food categories focus on the extremes of the low-calorie/high-calorie food axis. Although OLAF was specifically compiled to be used in studies on the affective and appetitive effects of food, it contains no food images with negative valence. To remedy the lack of negative valence images and to provide affective anchors, OLAF was extended with 36 non-food images from the IAPS (12 from each of the three valence categories pleasant, neutral, unpleasant) that cover the full valence-arousal space.

The Full4Health Image Collection (F4H, see: http://nutritionalneuroscience.eu) contains 228 images of Western food types of different caloric content, together with normative ratings (by adults from Greece, the Netherlands and Scotland, and by children from Germany, Hungary and Sweden) on recognizability, liking, healthiness and perceived number of calories. In addition, F4H also includes images of 73 non-food items. The images have a high resolution (3872×2592 pixels) and were registered according to a standardized photographing protocol (Charbonnier et al., 2016). F4H has been designed for health-related studies in which (perceived) caloric content is of interest and contains no food pictures with negative valence.

Food image databases for automatic recognition studies

Table 6-2 provides an overview of publicly available food image databases for automatic image recognition studies.

The Pittsburgh Fast-Food Image Dataset (PFID: (Chen et al., 2009); see: http://pfid.rit.albany.edu) contains still images, stereo pairs, 360-degrees videos and videos of Western fast-food and eating events, acquired in both restaurant environments and laboratory settings. The dataset represents 101 different foods with information on caloric content and is primarily intended for research on automated visual food recognition for dietary assessment. Although the images are registered at a high resolution (2592×1944 pixels), the products that are shown occupy only a small region of the image, while the luminance, structure and shadowing of the background vary largely across the image set (due to undulations in the grey cloth in the background). Since the database only contains images of fast-food, it covers only a small part of the valence-arousal space and is therefore not suitable for systematically studying the emotional impact of food. Also, the database is significantly (Western) cultural specific, implying the previously mentioned cross-cultural restrictions for this dataset as well.

The NU FOOD 360x10 database ((Takahashi et al., 2017); see: http://www.murase.is.i.nagoya-u.ac.jp/nufood) is a small food image database containing images of 10 different types of food, each shot at three elevation angles (30, 60 and 90 degrees) from 12 different angles (30 degree spacing). Six of the 10 foods are typically Asian (Sashimi, Curry and rice, Eel rice-bowl, Tempura rice-bowl, Fried pork rice-bowl, Tuna rice-bowl), while the remaining four represent Western food (Beef stew, Hamburger steak, Cheese burger, and Fish burger). The food categories were selected considering the variation of the appearance in both color and shape. However, for reasons of convenience and reproducibility plastic food samples were used instead of real ones. This may degrade the perceived naturalness of the images.

Table 6-2. Overview of food image databases for automatic recognition studies.

Databases	**Coverage of affective space**	**Recording methods**	**Cuisines**	**Availability of normative (and demographic) data**	**Remarks**
PFID (Chen et al. 2009)	A small part of valence and arousal space	Not standardized	Mainly Western	Not available	• High resolution (2592×1944 pixels) • Collected by the authors • No fixed background
NU FOOD (Takahashi et al. 2017)	Mainly positive valence	Standardized	Some Asian, Some Western	Not available	• Only 10 different cuisines (6 Asian and 4 Western cuisines) • No resolution specified
ChineseFoodNet (Chen et al. 2017)	Mainly positive valence	Not standardized	Only Chinese	Not available	• Variable resolution • Collected from Internet (185,628 images) • No fixed background
UNICT Food Dataset 889 (Farinella et al. 2015)	Mainly positive valence	Not standardized	Italian, English, Thailand, Indian, Japanese etc.	Not available	• Variable resolution • Collected with smartphones (3,583 images) • No fixed background
UEC-Food 100 (Matsuda et al. 2012) UEC-Food 256 (Kawano and Yanai 2015)	Mainly positive valence	Not standardized	France, Italy, US, China, Thailand, Vietnam, Japan, Indonesia, etc.	Not available	• Variable resolution • Collected from Internet • No fixed background
UPMCFOOD-101 (Wang et al. 2015) ETHZFOOD-101 (Bossard et al. 2014)	Mainly positive valence	Not standardized	More than 101 international food categories	Not available	• Variable resolution • Collected from Internet • No fixed background
VIREO-172 (Chen and Ngo 2016)	Mainly positive valence	Not standardized	Only Chinese	Not available	• Variable resolution • Collected from Internet (110,241 images) • No fixed background

The ChineseFoodNet ((Chen et al., 2017); see: https://sites.google.com/view/chinesefoodnet/) contains over 185,628 images of 208 Chinese food categories. The images in this database are collected through the internet and taken in real world under unconstrained conditions. The database is intended for the development and training of automatic food recognition algorithms.

The UNICT Food Dataset 889 (UNICT-FD889, (Farinella et al., 2015): see: http://iplab.dmi.unict.it/UNICT-FD889) contains 3,583 images of 889 distinct real meals of different nationalities (e.g., Italian, English, Thai, Indian, Japanese, etc.). The images are acquired with smartphones both with and without flash. Although it is an extended, cross-cultural database, images are not standardized and provided with emotional scores of e.g. valence and arousal. Additionally the technical quality of the presented images consequently fluctuates. This is most likely by design as the database is intended for the development of automatic image retrieval algorithms.

The UEC-Food100 (Matsuda et al., 2012) and UEC-Food256 ((Kawano and Yanai, 2015): see: http://foodcam.mobi/dataset.html) are both Japanese food image datasets, containing 100 and 256 food categories respectively, from various countries such as France, Italy, US, China, Thailand, Vietnam, Japan and Indonesia. The dataset was compiled to develop algorithms that automatically retrieve food images from the internet. The images have widely varying backgrounds (e.g. different compositions and lighting of plates etc.), implying that this has limited value for human neurophysiological food-related studies.

The UPMCFOOD-101 (Wang et al., 2015) and ETHZFOOD-101 ((Bossard et al., 2014); see: https://www.vision.ee.ethz.ch/datasets_extra/food-101/) datasets are twin datasets with the same 101 international food categories but different real-world images, all collected through the internet. The images of UPMCFOOD-101 are annotated with recipe information, and the images of ETHZ-FOOD-101 are selfies. The datasets were compiled to develop automatic systems for recipe recognition, an exercise that requires significantly other pictorial features than applications that intent to evoke discriminative, though well-defined, emotional responses.

VIREO-172 ((Chen and Ngo, 2016); see: http://vireo.cs.cityu.edu.hk/VireoFood172/) is a dataset containing 110,241 images of popular Chinese dishes from 172 categories, annotated with 353 ingredient labels. The images are retrieved from the internet and have widely varying backgrounds, implying the associated diversity in technical quality. Like some previously mentioned databases, this database is intended to develop automatic cooking recipe retrieval algorithms with ingredient recognition.

The CROCUFID database

Recording material and protocol

The CROCUFID images were created following the photographing protocol presented by Charbonnier et al. (2016), as shown in Figure 6-1. The images were taken with a Canon EOS 1300D high-resolution digital single lens mirror reflex camera that was mounted on a tripod and stored both in RAW (CR2) and JPEG format. The focal length used was 34.0 mm, the shutter speed was

1/50 s and the aperture value was 4.5 for each picture. In two pictures (image number 20 and 88) a blue plate was used to enable the study of color contrast effects (we plan to extend the database with more pictures of food arranged on plates with different colors in the future). Consequently, all other (food and non-food) items were placed on a white IKEA plate (type FÄRGRIK: www.ikea.com) with a diameter of 27 cm. The plate itself was placed in a 38×38×38 cm Foldio2 photo studio (a cubic photo tent made from white plastic sheets that soften and reflect the light from two LED strips above and below: see http://orangemonkie.com/foldio2). The camera was placed on a tripod. The Foldio2 studio and the camera were both rigidly attached to a common baseplate. The optical center of the camera was 39.5 cm from the center of the Foldio2 studio and 38 cm above the base plane. The viewing angle of the lens was approximately 45^o downwards, which resembles the viewing of a plate of food on a table during mealtime. A thin (5 mm thickness) grey foam board with a circular hole with the same diameter as the base of the IKEA plate was placed inside the Foldio2 studio, to ensure that the plate could easily be replaced in the same position for each image registration. The placement of the plate was checked by projecting a preregistered reference image of an empty plate over the actual plate using the 'Show Overlay' function of the camera. The foam board had a light grey color to ensure sufficient luminance contrast of the plate with its background, which enables automatic digital background correction (e.g., using Adobe Photoshop or Matlab). At the start of each food-registration session a picture of an X-Rite ColorChecker Passport (http://xritephoto.com) was made to enable post-registration white-balance correction and reproduction of the original colors in each image (using Adobe Photoshop or the accompanying ColorChecker Camera Calibration software). The images were partly registered in the Netherlands (N=426; images 1-357 and 565-633) and partly in Japan (N=414; images 358-564 and 634-840).

Figure 6-1. Standardized photographing protocol set-up (see: (Charbonnier et al., 2016)). A laptop (left) was used to control the camera (middle) settings and take the pictures of the plate with food in the Foldio2 photo studio (right).

The plate's background was standardized in all images as follows. First, a binary mask image was created in Photoshop by segmenting the reference image of the empty plate into a plate and background area followed by thresholding. Then, this mask image was used in Matlab 2018b (http://www.mathworks.com) to segment the plate from each of the 840 images and combine it with the segmented background from the reference image. Some smoothing was applied to the edges of the plate to prevent abrupt (noticeable) luminance transitions. In addition, the luminance of the (visible part of the) IKEA plate was equalized throughout the image set.

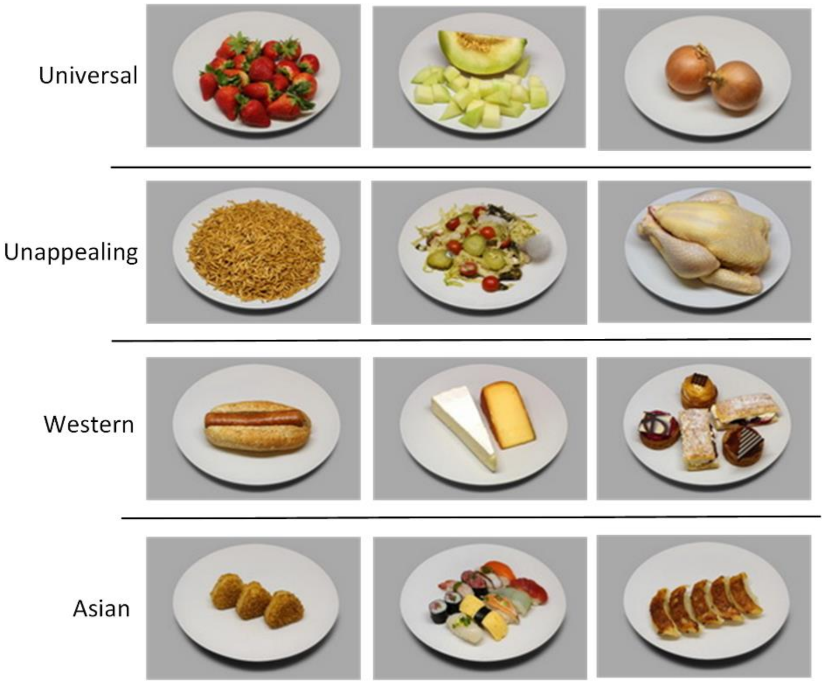


Figure 6-2. Representative images of four typical food categories (Universal, Unappealing, Western, Asian).

Figure 6-3. Representative images of three typical non-food categories (objects related or unrelated to food, flowers).

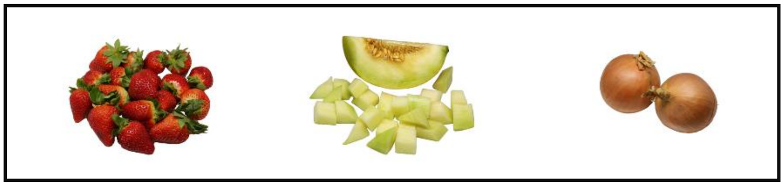


Figure 6-4. Some examples of item-only food images.

The CROCUFID database currently comprises 840 images: 675 food pictures (for some examples see: Figure 6-2) and 165 non-food pictures (Figure 6-3). Detailed metadata are provided for the first 479 food images. All images in CROCUFID are high-quality standardized 8 bits color pictures with a resolution of 5184 x 3456 pixels. For the validation studies reported in this paper the images were reduced in size to 1037 x 691 pixels. A corresponding set of item-only (food or non-food) images representing the displayed items on a transparent background (for some examples see Figure 6-4) was created by using the PhotoScissors background removal tool (www.photoscissors.com) for the initial separation of the food from the background, followed by the magic wand and color range selection tools in Adobe Photoshop to remove small remaining background sections inside the convex hull of the food area. These transparent item-only images can for instance be used to study the effects of background characteristics (e.g., plate size, color, texture, etc., see e.g. (Piqueras-Fiszman et al., 2012; Van Ittersum and Wansink, 2012; Piqueras-Fiszman et al., 2013; Stewart and Goss, 2013)) and plating arrangements (e.g., centered on a plate or off-center: (Michel et al., 2015a; Velasco et al., 2016)) by simply superimposing them onto images of different plates or even video clips of dynamic background textures (as in (Toet, 2016)). After thresholding and binarization the item-only PNG images can also serve as masks to restrict the calculation of computational measures to the area of the item on the plate.

Image features

Objective features like visual texture (Lucassen et al., 2011), complexity (Forsythe et al., 2011; Marin and Leder, 2013) and colorfulness (Cano et al., 2009) influence affective observer responses (measures of valence and arousal) to visual scenes in general (e.g., natural textures: (Toet et al., 2012)). Since our first sensory contact with food is typically through our eyes (van der Laan et al., 2011; Spence et al., 2015) it is not surprising that these visual cues also affect our responses to food (Zellner et al., 2010; Wadhera and Capaldi-Phillips, 2014). It has for instance been found that our appraisal and consumption of food depends on visual impressions like perceived regularity or randomness (Zellner et al., 2011; Zampollo et al., 2012), complexity (Mielby et al., 2012), spatial layout (Zampollo et al., 2012; Michel et al., 2015a; Szocs and Lefebvre, 2017), area of the plate covered (Van Ittersum and Wansink, 2012), and color content (Zellner et al., 2010; Zampollo et al., 2012; Piqueras-Fiszman and Spence, 2014; Spence, 2015b; Foroni et al., 2016; König and Renner, 2018). We therefore complemented the dataset with several computational measures that evaluate visual features (derived from the image luminance and chrominance distribution) that are known (or a priori likely) to influence affective image appraisal. These measures allow the user to select CROCUFID images based on their physical properties.

The first seven measures characterize the perceived image texture (i.e., the low-level spatial arrangement of color or intensities in an image). Five of these measures are only defined for grayscale images. To enable the computation of these measures the color images were therefore first converted to grayscale images with the MATLAB *rgb2gray* function before calculating these texture measures.

Entropy (S) is a statistical measure that characterizes the degree of randomness of the input image texture: entropy is 0 if all pixels have the same intensity value. Entropy was calculated with the standard MATLAB function *entropy*.

Power (P) is the average of the power spectral density over the image support (in decibels), computed from the discrete 2D Fourier transform of the image. This measure reflects the mean variations in image intensity.

The remaining three grayscale image texture measures (contrast, energy and homogeneity) were computed from the Grey Level Co-occurrence Matrix (GLCM, which is a classic technique used for image texture analysis and classification: (Haralick et al., 1976)) using the MATLAB function *graycoprops*.

Contrast (C) measures the intensity contrast between a pixel and its neighbors averaged over the whole image and is equal to 0 for a constant valued image (Corchs et al., 2016).

Energy (E) is the sum of squared elements in the GLCM and is equal to 1 for a constant valued image (Corchs et al., 2016).

Homogeneity (H) measures the closeness of the distribution of elements in the GLCM with respect to the GLCM diagonal and is equal to 1 for a diagonal GLCM (Corchs et al., 2016).

We also computed two measures that describe the texture of color images. These measures are based on the Pyramid Histogram of Oriented Gradients (PHOG) image representation that was originally developed for object recognition and classification (Bosch et al., 2007) and have been used to characterize the aesthetics and liking of images and artworks (Redies et al., 2012; Braun et al., 2013; Hayn-Leichsenring et al., 2017). The PHOG descriptors are global feature vectors based on a pyramidal subdivision of an image into sub-images, for which Histograms of Oriented Gradients (HOG: (Dalal and Triggs, 2005)) are computed.

Self-similarity(SS) is computed using the Histogram Intersection Kernel (HIK: (Barla et al., 2002)) to determine the similarity between Hog features at the individual levels of the PHOG (for details see (Redies et al., 2012; Braun et al., 2013)). Images of natural (growth) patterns typically have a highly self-similar (fractal) structure, whereas artificial (man-made) structures typically have a low self-similarity.

Anisotropy (AN) describes how the gradient strength varies across the orientations in an image. Low anisotropy means that the strengths of the orientations are uniform across orientations and high anisotropy means that orientations differ in their overall prominence (Redies et al., 2012).

The next six measures quantify the structural image complexity. The complexity of an image depends on the number of its structural components, their heterogeneity, (e.g., a single shape repeated vs. multiple distinct shapes), their regularity (e.g., simple polygons vs. more abstract shapes) and the regularity of the arrangement of elements (e.g., symmetry, distribution characteristics; see Figure 1 of (Berlyne, 1958)).

The *Compression Ratio* (CR) between the original (uncompressed) and (JPEG or GIF format) compressed file sizes is a computational measure that is positively correlated with ratings of subjective image complexity (Marin and Leder, 2013). The file size of a digitized image is a measure of its structural information content (Donderi, 2006). Compression algorithms use image redundancy or predictability to reduce the file size, such that more complex (or less predictable) images need more elements. Using the lossless JPEG compression mode of the MATLAB *imwrite* function we computed the JPEG based compression ratio (CRjpeg) which has been shown to be a reliable measure of subjective complexity across various image domains (Marin and Leder, 2013). We also computed the GIF based compression ratio (CRgif) using Adobe Photoshop (settings: palette local selective, colors 256, forced black-white colors, no transparency, dither diffusion 75%, exact colors and normal order of lines, see: (Marin and Leder, 2013)).

Feature Congestion (FC) is a visual clutter measure that implicitly captures the notion of spatial disorder by computing a weighted average of the local feature (color, orientation and luminance) contrast covariance over multiple (typically three) spatial scales (Rosenholtz et al., 2007). Larger FC values correspond to higher levels of visual clutter. The FC measure was calculated using the MATLAB code provided by Rosenholtz et al. (2007) (at http://www.mit.edu/~yzli/clutter.htm).

Subband Entropy (SE) is a clutter measure that encodes the image information content (or redundancy) by computing a weighted sum of the entropies of the luminance and chrominance image subbands (Rosenholtz et al., 2007). Larger SE values correspond to higher levels of visual clutter. The SE measure was calculated using the MATLAB code provided by Rosenholtz et al. (2007) (at http://www.mit.edu/~yzli/clutter.htm).

The *Number of Proto-Objects* (NPO) is the number of image segments or super-pixels with similar intensity, color and gradient orientation features (the proto-objects: (Yu et al., 2014)). Larger NPO values correspond to higher levels of visual clutter. The NPO measure was calculated using the MATLAB code provided by the authors (at http://www.chenpingyu.org/projects/proto-objects.html).

The *Mean Information Gain* (MIG) is defined as the difference between the spatial heterogeneity (i.e., the joint entropy among neighboring pixels) and the non-spatial heterogeneity (i.e., the probability of observing a pixel value independently of its location in the image) of an image (Andrienko et al., 2000; Proulx and Parrott, 2008). The MIG does not require knowledge of the `maximal' entropy of the image (which is sometimes hard to compute or even to define), and accounts for the inherent spatial correlations. The MIG increases monotonously with spatial randomness and ranges over 0-1: MIG=0 for uniform patterns and MIG=1 for random patterns. The MIG index is a well-known complexity measure in statistical physics (Wackerbauer et al., 1994) and has successfully been used to quantify the complexity of two-dimensional patterns (Andrienko et al., 2000) and ecological habitats (Proulx and Parrott, 2008). The images were transformed to HSV format and the pixel value range was normalized from 0-10 before calculating MIG values independently for the color (Hue: MIGh), chroma (Saturation: MIGs) and intensity (Value: MIGv) components of each image. The MIG measure was calculated using the MATLAB code provided by Proulx and Parrott (2008) at http://complexity.ok.ubc.ca/projects/measuring-complexity.

The *Mean Gradient Strength* (MGS) or mean edge strength is a valid measure of subjective image complexity (Braun et al., 2013) and is based on the observation that the subjectively perceived level of image complexity increases with its number of edges. To compute the MGS we first transformed the images from RGB to Lab format (using the MATLAB function *rgb2lab*). Then the MATLAB function *gradient* was applied to each of the three image channels and a single gradient image was obtained by taking the pixelwise maximum of the three individual gradient images. Finally, the mean of the resulting gradient image was adopted as an overall measure of image complexity (Braun et al., 2013).

The following five measures characterize the image color distribution.

The *Number of Colors* (NC) represents the number of distinct colors in the RGB image. For each image this number was obtained as the size of the color map resulting from the application of the MATLAB *rgb2ind* function (with the minimum variance quantization and dithering options) to each original RGB image.

Colorfulness is the sensation that an image appears to be more or less chromatic. Local colorfulness has been defined as a linear combination of the mean and standard deviation of the local chrominance values in color opponent space (Hasler and Süsstrunk, 2003). Note that colorfulness is not strictly related to the numbers of colors: an image can be more colorful even when its contains less different colors (Palus, 2005). A global image Colorfulness (CF) metric was computed as the mean value of the local colorfulness over a set of subwindows covering the entire image support (Hasler and Süsstrunk, 2003). CF varies from 0 (grayscale image) to 1 (most colorful image). CF was computed using the MATLAB code obtained from https://gist.github.com/zabela/8539116.

The proportional contribution of the Red (R), Green (G) and Blue (B) color channels was computed from the transparent PNG images to characterize the overall color of the (food or non-food) items (as in (Blechert et al., 2014b)).

Finally, the *Fraction* of the *Plate Covered* (FPC) by the (food or non-food) item on the plate was calculated for each image as follows. The total number of pixels over the area of the plate was calculated as the number of non-zero pixels in binary mask image of the empty plate. The total number of pixels covered by the item on the plate was calculated as the number of non-zero pixels in the binary image resulting after thresholding and binarization of the corresponding transparent PNG version of the image (note that this operation yields a binary mask of the area actually covered by the item). The FPC was then obtained as the ratio of both these numbers (i.e., the number of pixels representing the item divided by the number of pixels representing the plate). Note that the FPC can also be adopted as a measure for stimulus size (see: (Foroni et al., 2013)) since the plate has the same size in all images.

CROCUFID availability and use

The CROCUFID database is publicly available from the Open Science Framework repository (OSF) at https://osf.io/5jtqx with DOI 10.17605/OSF.IO/5JTQX) under the CC-By Attribution 4.0 International license. Use is only allowed after complying with the following two conditions: (1) a credit line in publications and presentations reading: "*Standardized images were taken from the CROCUFID database available from the OSF repository at https://osf.io/5jtqx*", and (2) a citation to the current article in any publication.

The database includes: the CROCUFID pictures; contact sheets providing a concise overview of all images in the dataset; documents (both in Adobe Acrobat and in Powerpoint) containing screen shots of the entire observer validation experiment; an SPSS file with the raw observer data (each participant's code and demographic data, followed by valence, arousal, perceived healthiness and desire-to-eat ratings); an Excel file listing for each image some classifiers, the computational measures, and the mean and standard deviations of the observer ratings for the food images. The full-scene images (showing the plated items on a grey background) in the CROCUFID database are stored in JPEG format, while the item-only images (isolated items displayed on a transparent background) are in PNG format. The original RAW images and the calibration images showing the X-Rite ColorChecker Passport are available (in CR2 or DNG format) from the authors on request. The CROCUFID image database complements the F4H image collection (http://nutritionalneuroscience.eu) since both were registered using the same protocol (Charbonnier et al., 2016).

Cross-cultural evaluation study

In our global economy cross-cultural research is becoming increasingly relevant in sensory and consumer science (Meiselman, 2013a). Culture is one of the main factors determining human food-related behavior (Rozin, 1988). Internet-based research provides almost instantaneous world-wide access to large consumer samples, enabling researchers to conduct cross-cultural studies at relatively low-cost and in short time frames (Slater and Yani-de-Soriano, 2010; Ares, 2018). Online cross-cultural studies using images of food have for instance successfully been performed to study the effects of spatial arrangement and color composition of meals on food preference (Zampollo et al., 2012). Also, we recently used CROCUFID images in an online cross-cultural validation study of a new affective self-report tool (Kaneko et al., 2018c). The item-only images in CROCUFID enable cross-cultural (online) HCI studies on the effects of environment and context on food experience (see previous section). The CROCUFID food image database is a useful tool for food-related cross-cultural research since it contains food pictures from different cuisines. Therefore, we conducted a first cross-cultural study as outlined below.

Methods

Participants

Three groups completed an online anonymous survey to provide normative data for the food images in CROCUFID. The total number of participants was 805.

Two groups consisted of English speaking participants that were recruited through the Prolific survey site (www.prolific.ac). The text of the experiment posted on Prolific was in English. The first group (UK) comprised 266 participants from the UK. The second group (US) comprised 275 US participants. The third group (JP) consisted of Japanese speaking participants that were recruited through the Crowdworks survey site (https://crowdworks.jp). The text of the experiment was translated into Japanese for this sample. The validity of the translation was checked using the back-translation technique (Sperber, 2004). The third group comprised 264 participants. Exclusion criteria for all participants were color blindness and age (younger than 18 or older than 70).

All participants signed an informed consent form before taking part in the study and received a small financial compensation after completing the study. The experimental protocol was reviewed and approved by the TNO Ethics Committee (Ethical Approval Ref: 2017-016-EM) and was in accordance with the Helsinki Declaration of 1975, as revised in 2013 (World Medical Association, 2013).

Stimuli

The compilation of the CROCUFID database continued in parallel during this entire study. For the observer validation experiments reported in this study we used only the 479 food images that were already available at the start of the experiments. The remaining 361 images in CROCUFID were registered during and after the observer experiments. In the present study the stimuli were classified into four categories: a Universal food category (consisting of universally well-known fruits and vegetables etc.: N = 110), a typical Western food category (consisting of sandwiches, cookies, cheeses, cakes etc.: N = 119), a typical Asian food category (consisting of sushi, sashimi, rice-bowl, noodles etc.: N = 209) and an Unappealing food category (consisting of unfamiliar, rotten, molded or contaminated food: N = 41). Unfamiliar food was classified as unappealing since people typically evaluate (e.g., liking and willingness to try) novel foods more negative compared to familiar foods (Raudenbush and Frank, 1999).

Demographics and state variables

Participants were asked to report their personal characteristics. They provided gender, age, height, weight, nationality and eating habits, such as dieting and food allergies. Participants were not requested to refrain from eating prior to taking part in the experiment, but their psychophysiological state was registered by asking them how hungry and thirsty they were feeling at the time of the experiment and how much time had passed since their last food consumption. Their state of hunger and thirst were measured with visual analog scales (VAS) labeled from "*very hungry/thirsty*" to "*not hungry/thirsty at all*". The VAS was displayed as a solid horizontal bar, and participants responded by clicking with the mouse on the appropriate location of the line. Responses were analyzed by converting distances along the bar to a scale ranging from 0 to 100 although this was not explicitly displayed to the participants. BMI was also calculated based on participants' reported height and weight.

Subjective image measures

For all images, valence, arousal, perceived healthiness and desire-to-eat were measured using a VAS ranging from 0 to 100 (again, this was not explicitly displayed to the participants). This is a valid method to measure food elicited affective responses (Flint et al., 2000) that has previously been used to validate food image databases (Foroni et al., 2013; Blechert et al., 2014b). Valence expresses the pleasantness of an image. The question was: "*How pleasant is the item presented in* *the image*?" and the extremes of the scale were labelled as "*Very unpleasant*" (0) and "*Very pleasant*" (100). The instructions for the participants explained the concept of valence by presenting some related terms for both "*Very unpleasant*" (bad, disliked, disgusting, unsatisfied, annoyed) and "*Very pleasant*" (good, liked, delicious, satisfied, pleased). Arousal measures the excitement (or intensity) that was experienced while viewing a food picture. The question was: "*How arousing is the item presented in the image*?" and the extremes of the scale were labeled as "*Not at all*" and "*Extremely*". The instructions for the participants explained the concept of arousal by presenting some related terms for both the lower (calming, relaxing) and upper (stimulating, energizing) ends of the scale. Perceived healthiness was assessed with the question: "*How healthy do you think the item presented in this image is*?" and the extremes of the scale were labeled as "*Very unhealthy*" and "*Very healthy*". Desire-to-eat was assessed with the question: "*How much would you like to eat this food right now if it was in front of you*?" and the extremes of the scale were labeled as "*Not at all*" and "*Extremely*". The degree of recognition of the food images was also assessed by asking the question: "*Have you ever eaten the product in this image*?" that could be rated on a trichotomous scale: "*Yes*", "*No, but I know what it is (Know)*" or "*No, I've never seen it before (No)*".

Analysis

Normative observer data for the food images in CROCUFID were obtained through an anonymous online survey. As participants could not be expected to reliably rate all 479 food images, each participant rated only a subset of 60 images to avoid fatigue and early dropout. On average, each image was rated by approximately 100 participants including all participants from three different countries (UK, US, and JP).

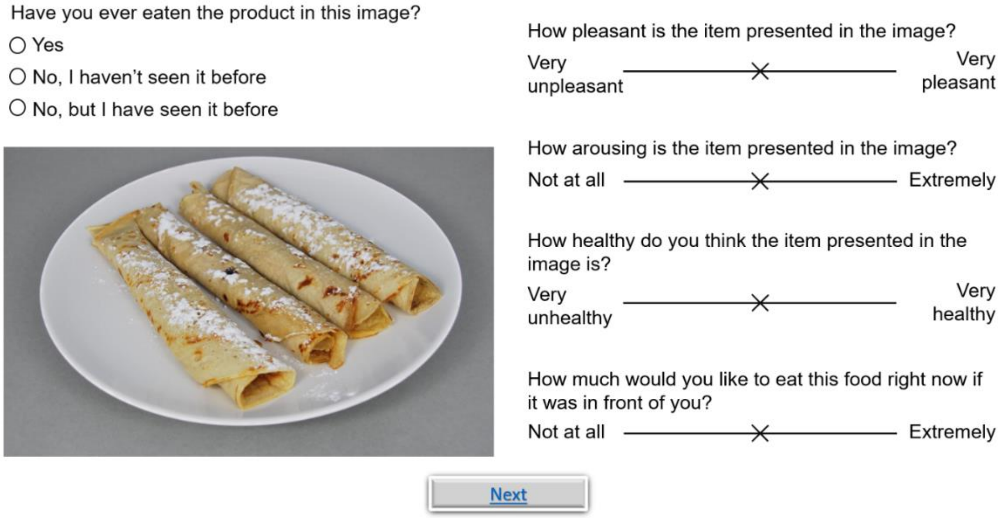
Data was collected using Perl scripts (www.perl.com) and analyzed with IBM SPSS Statistics 23 (www.ibm.com). Exploratory analyses were conducted between all demographics (i.e., age, gender, nationality, height, weight, diet, allergies), state variables (hunger, thirstiness, and time since last food intake) and subjective self-report ratings (valence, arousal, perceived healthiness and desire-to-eat) of four classified food image categories (Universal foods, Unappealing foods, typical Western foods, and typical Asian foods) from three different nationalities (UK, US, and JP). In addition, the degree of recognition of four different food image categories was calculated from UK, US, and JP participants. Participants who answered "*Yes*" or "*No, but I know what it is (Know)*" were categorized as 'recognizing' participants, and participants who answered "*No, I've never seen it before*" were categorized as 'non-recognizing' participants.

Intraclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated based on a mean-rating ($k = 3$), absolute-agreement, 2-way mixed-effects model (Shrout and Fleiss, 1979; Koo and Li, 2016). ICC values less than .50 are indicative of poor reliability, values between .50 and .75 indicate moderate reliability, values between .75 and .90 indicate good reliability, while values greater than .90 indicate excellent reliability (Koo and Li, 2016).

Procedures

Participants took part in an anonymous online survey. Although internet surveys typically provide less control over the experimental conditions, they typically yield similar results as lab studies (e.g., (Gosling et al., 2004; Woods et al., 2015; Majima et al., 2017)) while they limiting several disadvantages associated with central location studies.

The experiment was programmed in the Java script language, and the survey itself was hosted on a web server. The time stamps of the different events (onset stimulus presentation, response clicks) and the display size and operating system of the participants were logged. This enabled us to check that participants did indeed view the stimuli on larger displays and not on mobile devices with low resolution screens. The resolution of the devices used by the participants in this study varied between 1280x720 and 3440x1440 (the average resolution was 1538x904 pixels across participants, with standard deviations of 330x165 pixels). We could not verify if the browser window was indeed maximized.



Have you ever eaten the product in this image?

○ Yes

○ No, I haven't seen it before

○ No, but I have seen it before

How pleasant is the item presented in the image?

Very unpleasant — Very pleasant

How arousing is the item presented in the image?

Not at all — Extremely

How healthy do you think the item presented in the image is?

Very unhealthy — Very healthy

How much would you like to eat this food right now if it was in front of you?

Not at all — Extremely

Figure 6-5. Screenshot of the display during the image rating task.

The survey commenced by presenting general information about the experiment and thanking participants for their interest. Then, the participants were informed that they would see 60 different food images during the experiment, and they were instructed to rate their first impression of each image without worrying about calories. It was emphasized that there were no correct or incorrect answers, and that it was important to respond seriously and intuitively. Subsequently, the participants signed an informed consent by clicking "*I agree to participate in this study*", affirming that they were at least 18 years old and voluntarily participated in the study. The survey then continued with an assessment of the demographics and the current physical state of the participants. Next, the participants were shown the VAS response tools together with an explanation how these could be used to report their ratings (valence, arousal, perceived healthiness and desire to eat) for each image. Then, they performed two practice trials to further familiarize them with the use of the VAS tools. Immediately after these practice trials, the actual experiment started. During the actual experiment the participants rated 60 pseudo-randomly selected food images. The selection procedure was such that 14 images were randomly selected from the 110 Universal images, 5 from the 41 Unappealing/spoiled images, 26 from the 209 typical Asian images, and 15 from the 119 typical Western images. This selection procedure ensured that all images were rated by approximately the same number of participants and all participants viewed the same number of images from each category. The percentage of images from each category that each participant actually viewed was equal to the percentage of that particular category in the total stimulus set.

On each trial the screen displayed a food image together with the recognizability question and the four VAS tools to rate valence, arousal, perceived healthiness and desire-to-eat (for a screenshot of the experiment see: Figure 6-5). Since this experiment was an online study with a (small) financial compensation and without any personal contact or interaction with the participants, there might be issues with motivation and seriousness. Therefore, the experiment concluded with a validated seriousness check (Aust et al., 2013) asking participants the question: "*It would be very helpful if you could tell us at this point whether you have taken part seriously, so that we can use your answers for our scientific analysis, or whether you were just clicking through to take a look at the survey?*". Participants could select one of two answers: "*I have taken part seriously*" or "*I have just clicked through, please throw my data away*". As an extra incentive, the following sentence was added to the instructions that were presented at the start of the experiment: "*It's important for us that you are motivated and answer all questions seriously*". After completing the experiment, participants received a small financial compensation and were thanked for their participation. The whole session lasted about 20 minutes.

Results

There were no participants that (1) completed the study in a time span that was evidently too short in comparison to the estimated total survey duration (suggesting that they did not participate seriously) or that (2) responded negatively to the question about their seriousness were removed from further analysis. Hence, all data collected were included in the analysis.

Demographic information

The average age, hunger, thirst, time since last food intake and BMI (weight in kilograms divided by height in meters squared), and the number of each gender for the three different participant groups (UK, US, and JP) is summarized in Table 3. There was a significant difference between the average age of the UK and US participants. The average time since last food-intake was significantly shorter for JP participants than for UK and US participants, which is reflected in the result that JP participants were significantly less hungry than UK and US participants. There was no significant difference between three nationality groups regarding thirst.

Data reliability

To quantify the agreement between the mean subjective image ratings provided by participants from the three different countries (JP, UK and US), we computed the intraclass correlation coefficient (ICC) between each pair of countries, for each of the four subjective ratings (valence, arousal, desire-to-eat and perceived healthiness). As shown in Figure 6-6, there is excellent agreement (all ICC values are larger than .90) between the UK and US groups for each of the four subjective ratings. The JP group shows moderate (between .50 and .75) to good (between .75 and .90) agreement with the UK and US groups. Overall, there appears to be a trend that the JP group agrees more with the US group than with the UK group.

Table 6-3. The summary of participants' demographics from three different nationalities (United Kingdom (UK), United States (US), and Japan(JP)).

	UK	US	JP
Number of participants	266	275	264
Age	36.68 (± 11.26)	33.04 (± 11.38)	35.17 (± 8.95)
Gender male (female)	83 (183)	141 (134)	101 (163)
Hunger	42.04 (± 27.63)	41.34 (± 26.85)	31.52 (± 25.55)
Thirst	45.96 (± 24.39)	44.54 (± 24.13)	42.63 (± 22.31)
Time since last food-intake (hr.)	4.20 (± 4.59)	4.97 (± 4.66)	3.40 (± 2.33)
BMI	28.94 (± 12.54)	27.27 (± 9.37)	21.47 (± 3.18)

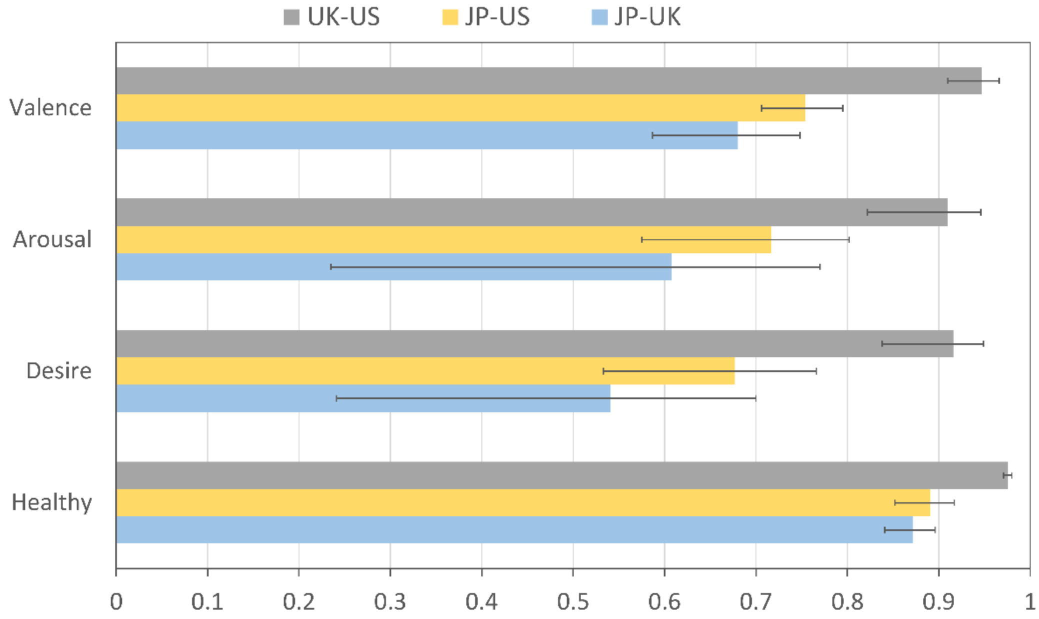
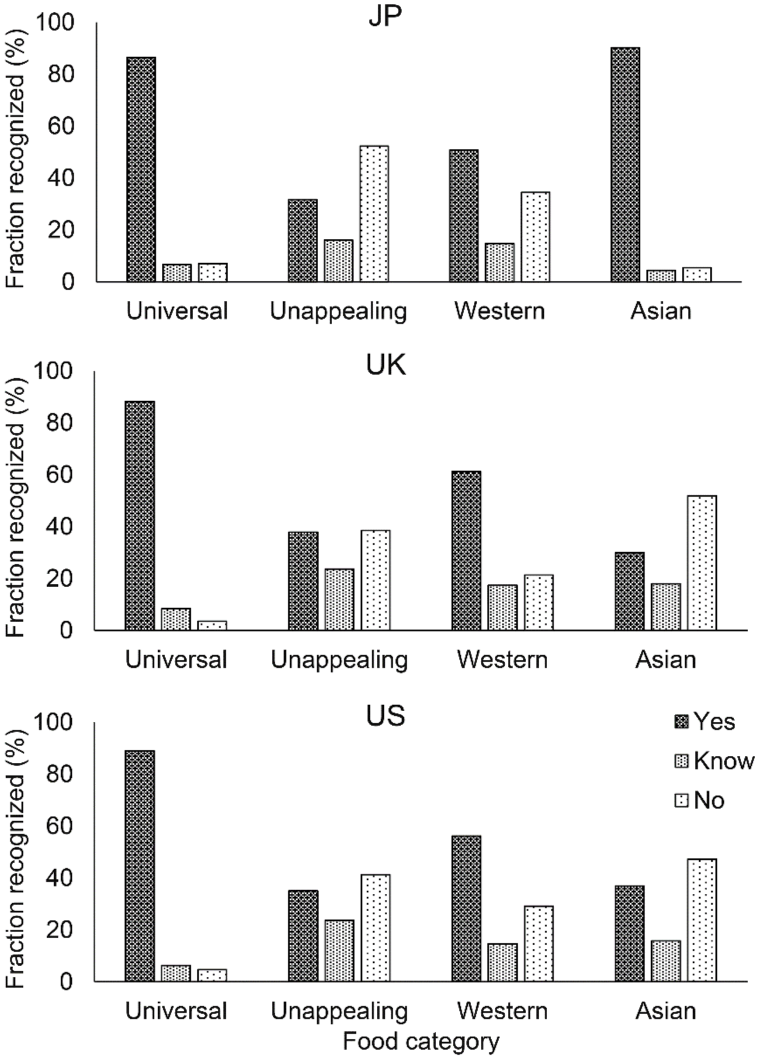


Figure 6-6. Intraclass correlation between the mean subjective (valence, arousal, desire-to-eat and perceived healthiness) ratings for each of the three different nations (United Kingdom, United States, and Japan). Error bars represent the 95% confidence intervals.

Validation ratings

For all 479 food images, we computed the average rating scores of valence, arousal, perceived healthiness, and desire-to-eat, both over all groups and within groups (UK, US, and JP). The results are provided as additional material with this paper. The food images were categorized in four different categories: Universal food (food that is globally available), Unappealing food (i.e., unfamiliar, rotten, molded or contaminated food), typical Western food, and typical Asian food.



For all three groups tested , the rating scores for Universal food images are higher than those for Unappealing food images.

Figure 6-7. The average degree of recognition of four categories of food images rated by JP participants, UK participants, and US participants.

Degree of recognition on food images across groups: We also evaluated the degree of recognition for each of the four categorized food image groups across the three different groups (UK, US, and JP) . Figure 6-7 shows for each population the average percentage who answered "*Yes*", "*No, but I know what it is (Know)*", and "*No, I've never seen it before (No)*" for each food image group.

Over all three populations, 94.9 % of participants recognized Universal food images on average (separately 96.6 % of UK, 95.3 % of US, and 93.0 % of JP). UK and US participants recognized significantly more Western than Asian food images. JP participants recognized most of the Asian food images (94.6 %), while only 48.0 % of UK and 52.7 % of US participants recognized them. There was a significant difference in the degree of recognition of Asian food images between US and JP participants, and between UK and JP participants. US participants recognized significantly more Asian images than UK participants.

Comparison of rating scores on Asian and Western food categories across groups (valence, arousal, perceived healthiness, and desire to eat): Next, we compared the average rating scores of valence, arousal, healthiness and desire-to-eat for Western (Figure 6-8A) and Asian (Figure 6-8B) food categories between UK and JP and between US and JP participants. US and UK participants rated Western food significantly higher on valence than JP participants. There was no significant difference between the rating scores on arousal and desire-to-eat for Western food and for each of the groups tested (US and UK and JP). JP participants rated Asian food images significantly higher than UK and US participants on all items.

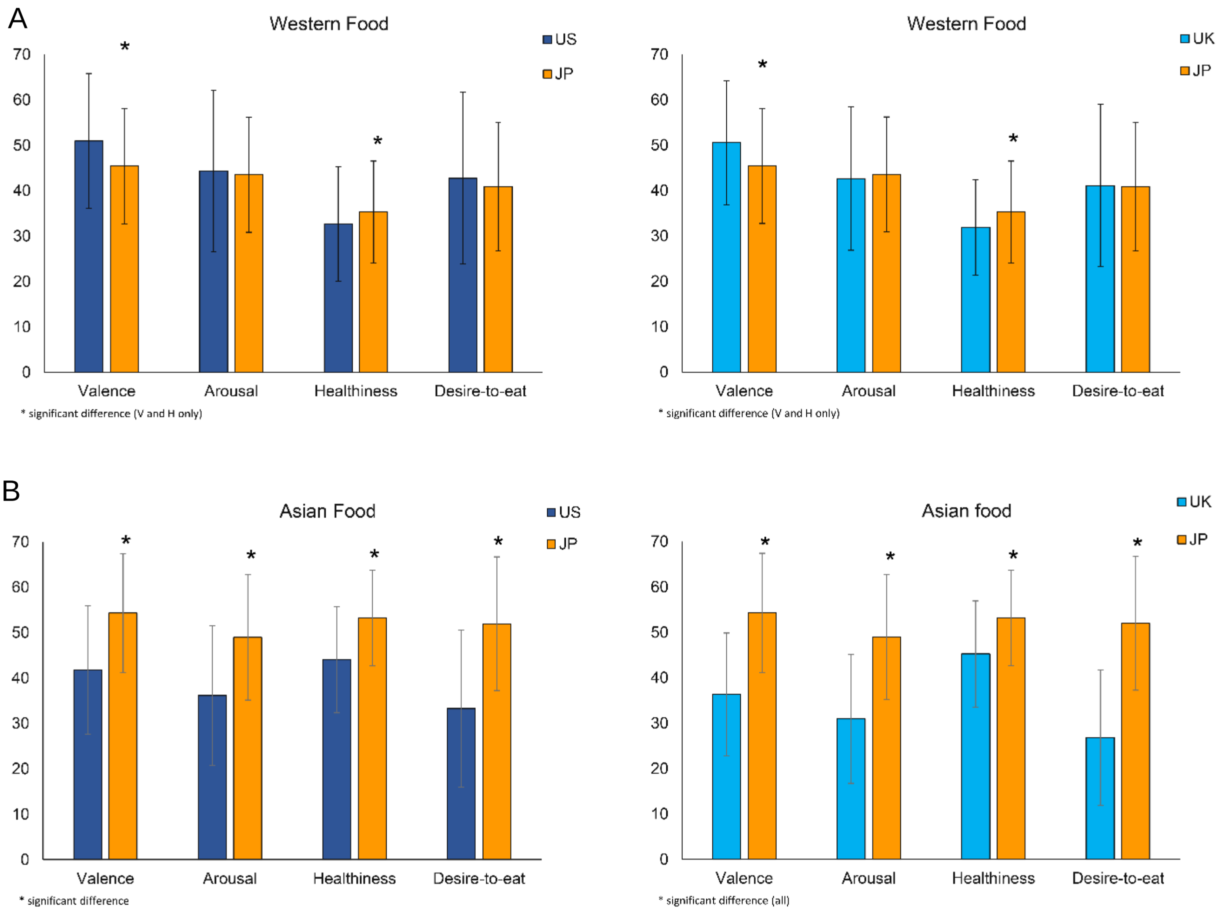


Figure 6-8. The comparison of the average rated scores of valence, arousal, healthiness and desire-to-eat on Western (A) and Asian (B) food category between UK and JP and between US and JP participants. *Indicates a significant difference between groups.

Computational image measures

The main purpose of providing computational image characteristics with CROCUFID is to allow the user to select CROCUFID images based on their physical properties. Since the selected features are known (or a priori likely) to influence image appraisal we also computed the correlation between the computational metrics and the mean observer ratings. Table 6-4 lists the significant Pearson correlations between the mean observer ratings (overall and for each of the three groups individually) and each of the computational image measures.

Although most correlations are small (between 0 and .3), there is still an appreciable number of medium sized correlations (between .3 and .5), particularly for the texture metrics AN and SS, and the color metrics NC and CF.

Image texture: The image texture metrics S, SS and AN appear to have the largest absolute overall correlation with mean observer ratings for valence, arousal and desire-to-eat. For the UK and US groups, S and SS are positively correlated with all four of the mean observer ratings, meaning that for these groups valence, arousal, perceived healthiness and desire-to-eat all increase with increasing randomness (S) and self-similarity (SS) of the food items. For the JP group, S is only weakly positively correlated with arousal, while SS is not significantly correlated with any of the mean ratings, suggesting that randomness and self-similarity do not strongly affect the appraisal of food items for this group. AN (which is inversely related to disorder) correlates negatively with each of the four mean ratings for the UK and US groups, while it correlates positively with each of the mean ratings for the JP group. This means that increasing variations in local orientations over the area of a food item decreases each of the four mean ratings for the UK and US groups, while it increases these ratings for the JP group.

Complexity: The complexity metrics CR_{jpeg}, CR_{gif} and FC appear to have the largest absolute positive overall correlation with mean observer ratings for valence, arousal and desire-to-eat. This correlation is positive for the UK and US groups, but negative for the JP group, meaning that perceived valence, arousal and desire-to-eat increase with increasing complexity for the UK and US groups, while they decrease with complexity for the JP group. NPO is the only complexity measure that systematically correlates positively with all four measures for all three groups. This means that all groups rate perceived healthiness higher when food items are composed of a larger number of distinguishable components.

Color: The color metrics NC and CF are positively associated with all four measures for all three groups. However, compared to the UK and US groups, this correlation is quite weak for the JP group. Hence, it seems that valence, arousal, perceived healthiness and desire-to-eat all increase with increasing number of colors and colorfulness of the food items, but more so for the UK and US groups than for the JP group.

Cultural differences: The results listed in Table 6-4 show that the relation between the mean subjective image ratings and each of the most predictive image metrics differs consistently between the JP groups and the UK and US groups: the variation of the mean (valence, arousal, desire-to-eat and perceived healthiness) ratings of the JP group with each of the texture, complexity and color metrics is consistently smaller or even opposite to those of the UK and US groups. A graphical overview of these results is provided in the Supplementary Material with this article.

Arousal, valence and desire-to-eat consistently increase with increasing structural food complexity (i.e., with increasing values of the texture, complexity and color metrics) for the UK and US groups (see: Table 6-4). On these three subjective measures, the JP group shows an opposite behavior to the UK and US groups for the texture measure AN and for each of the three complexity measures CR_{jpeg}, CR_{gif} and FC. Like the UK and US groups, the mean arousal ratings by the JP group also increase with increasing food item colorfulness or number of colors, though less strongly. Perceived healthiness increases with the number of colors (NC), distinguishable components (NPO) and randomness (S) of the food items, though less strongly for the JP group than for the UK and US groups.

Discussion

The CROCUFID database contains high-quality images that are registered according to standardized protocol, covering the full valence and arousal space. CROCUFID includes multiple cuisines and unfamiliar foods and provides normative and demographic data. The database is hosted in the OSF repository (https://osf.io/5jtqx) and is freely available under the CC-By Attribution 4.0 International license. We plan to extend this image collection with food images from different cuisines in addition to the Western and Asian foods that are currently included. Also, the CROCUFID image collection can easily be extended by linking it to similar data sets. Researchers who are interested in contributing to this effort are kindly invited to provide the authors with a link to their image dataset. The images should preferably be documented and registered according to the protocol presented by Charbonnier et al. (2016).

As expected, the present results of our cross-cultural study show that the JP group recognized significantly more Asian food compared to UK and US groups, while the UK and US groups recognized significantly more Western food than the JP groups, suggesting that Asian food images seem to have a strong cultural “bias”: for Western group (US and UK) it is hard to distinguish different food images. As also expected, JP group who has a higher degree of recognition of Asian food images rated significantly higher scores of valence, arousal, healthiness, and desire-to-eat than Western groups (US and UK). Regarding the rating scores on Western food images, Western groups (US and UK) rated significantly higher scores of valence than JP group. There was no significant difference in the rating scores of arousal and desire-to-eat between the western group and the JP group. The results of valence ratings on Asian and Western food by the JP group and the Western groups (UK and US) agree with the previous findings of Knaapila et al. (2017) that the familiarity of spice odors the participants have was positively associated with pleasantness.

Table 6-4. Pearson correlations between the mean observer ratings (overall and for each of the three groups individually) and the computational image measures.

		Computational measure																				
		Texture							**Complexity**									**Color**				
Mean observer rating		**S**	**P**	**C**	**E**	**H**	**SS**	**AN**	**CRjpeg**	**CRgif**	**FC**	**SE**	**NPO**	**MIGh**	**MIGs**	**MIGv**	**MGS**	**NC**	**CF**	**R**	**G**	**B**
Valence	**All**	**.238****			-.153**		**.201****	-.151**	.137**	.133**	.177**		.109*					**.268****	**.353****			
	UK	**.264****			-.179**	-.095*	**.300****	**-.329****	**.243****	**.226****	**.295****	.099*	.130**		.152**	.122**	.134**	**.290****	**.406****			
	US	**.273****			-.170**		**.256****	**-.253****	.200**	.194**	**.254****		.125**		.119**		.103*	**.278****	**.366****			
	JP		-.124**	-.133**		.118**		**.242****	-.117*	-.100*	-.124**	-.151**				-.149**	-.100*	.137**	.148**			
Arousal	**All**	**.266****			-.190**		**.219****	-.187**	.190**	.170**	**.204****		.158**		.144**		.122**	**.345****	**.375****			
	UK	**.260****			-.189**	-.119**	**.294****	**-.328****	**.258****	**.230****	**.288****	.120**	.167**		.179**	.148**	.159**	**.328****	**.402****			
	US	**.295****			-.199**	-.112*	**.257****	**-.269****	**.236****	**.214****	**.262****	.109*	.170**		.168**	.118**	.153**	**.344****	**.365****			
	JP	.133**			-.109*			.161**										**.238****	**.207****			
Healthiness	**All**	.200**	-.092*				.148**						**.312****					**.204****	.134**	-.119	.148**	
	UK	**.239****					.189**						**.317****					**.243****	.167**	-.125	.150**	
	US	**.233****					**.208****						**.314****					**.223****	.172**	-.115	.155**	
	JP		-.136**	-.124**		.160**		.189**	-.172**	-.132**	-.200**		**.255****		-.101*	-.170**	-.120**	.099*		-.098	.113*	
Desire-to-eat	**All**	**.244****			-.153**		**.214****	-.161**	.170**	.167**	**.209****		.105*		.109*			**.244****	**.353****			
	UK	**.275****		.097*	-.184**	-.141**	**.334****	**-.382****	**.300****	**.283****	**.352****	.144**	.122**		.186**	.179**	.176**	**.261****	**.410****			
	US	**.280****			-.188**	-.111*	**.267****	**-.282****	**.249****	**.237****	**.300****	.103*	.123**		.157**	.123**	.143**	**.259****	**.374****			
	JP		-.125**	-.137**		.124**		**.275****	-.128**	-.107*	-.132**	-.158**				-.155**	-.109*	.107*	.119**			

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Values above 0.2 are printed in bold. Only significant correlations are reported.

Fenko et al. (2015a) also found similar results using soy products, where familiar products were evaluated more positively than unfamiliar products with German participants.

Our current result that valence, arousal, desire-to-eat and perceived healthiness are positively associated with all texture measures (spatial disorder) for the UK and US groups and less strongly with some of them for the JP group agrees with the previous finding of Zampollo et al. (2012) that participants from the United States and Italy tend to prefer disorganized food presentations, while Japanese participants do not have a significant preference.

The present result that valence, arousal, desire-to-eat and perceived healthiness are positively associated with all measures of complexity for the UK and US groups agrees with the results of many previous studies that reported a positive association between visual complexity and affective (i.e., valence and arousal) ratings: complex stimuli are typically perceived as more pleasant and more arousing than less complex ones (for an overview see (Madan et al., 2018)). However, the JP group shows the opposite behavior: mean valence, arousal and desire-to-eat ratings are all negatively associated with complexity for this group.

Our finding that the number of identifiable components (NPO) is positively associated with perceived healthiness agrees with previous observations that (1) people prefer servings composed of multiple pieces to single-piece ones (Wadhera and Capaldi-Phillips, 2014), and (2) food cut in smaller pieces is preferred to larger chunks (Wadhera and Capaldi-Phillips, 2014), probably because segregating food into multiple components increases the perceived variety of the foods and is therefore perceived as more rewarding (Levitsky et al., 2012). It has also been suggested that perceived variety (e.g., through multiple colors or components) may be preferred in food since it typically delays the sensory-specific satiety (Wadhera and Capaldi-Phillips, 2014).

Our finding that a higher intensity of the green color channel is positively associated with perceived healthiness, while a higher intensity of the red color channel is negatively correlated with healthiness, agrees with previous reports that a higher intensity of the green channel was positively associated with lower concentrations of protein, fat and carbohydrates (Blechert et al., 2014b)) and with a lower (perceived) number of calories (Blechert et al., 2014b; Foroni et al., 2016), while a higher intensity of the red channel was positively associated with energy density (Foroni et al., 2016).

Finally, the result that the number of colors (NC) and colorfulness (CF) are positively associated with that the mean valence, arousal, desire-to-eat and perceived healthiness ratings agrees with previous reports in the literature that multicolored food is rated higher in attractiveness (Zellner et al., 2010) and pleasantness (Rolls et al., 1982) than single-colored food, and that colorfulness is typically positively associated with healthiness (König and Renner, 2018).

This study also has some limitations. First, since the validation rating internet survey was conducted during the construction of CROCUFID, not all food (non-food) images have been rated yet. Second, since all participants were randomly collected from the UK, US, and Japan, there is no information about their exact region of origin. Hence, additional validation survey is needed to complete the database and specify cross-cultural effects within each country.

In conclusion, CROCUFID currently contains 840 food images (675 food images and 165 non-food images), from different (Western and Asian) cuisines and with different degrees of valence.

CROCUFID also includes computational image characteristics regarding visual texture, complexity, and colorfulness, that may be used to select images for different applications (e.g., food research, automatic image interpretation, and VR/AR applications). The accompanying validation data are derived from a total of 805 participants (ranging in age from of 18-70) with different cultural backgrounds (UK, US, and JP). CROCUFID may also be used to conduct (neuro)physiological studies because all items are shown in the same background context and are also available as transparent overlays. Currently many different food cuisines are still lacking from CROCUFID. We hope to extend the collection with images of food from other cuisines. Researchers from different parts of the world are kindly invited to contribute to this effort by creating similar image sets that can be linked to this collection, so that CROCUFID will grow into a truly multicultural food database.

Acknowledgements

The authors thank prof. dr. Christoph Redies (Experimental Aesthetics Group, Institute of Anatomy, University of Jena School of Medicine, Jena University Hospital, Jena, Germany) for sharing his MATLAB software to calculate the PHOG based (self-similarity, complexity and anisotropy) measures.

Supplementary Material

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fpsyg. 2019.00058/full#supplementary-material.

Chapter 7

EmojiGrid: A 2D Pictorial Scale for the Assessment of Food Elicited Emotions

This chapter is published as:
Toet, A., Kaneko, D., Ushiama, S., Hoving, S., De Kruijf, I., Brouwer, A.M., Kallen, V., and Van Erp, J.B.F. (2018). EmojiGrid: A 2D pictorial scale for the assessment of food elicited emotions. Frontiers in psychology 9, 2396.

Abstract

Research on food-experience is typically challenged by the way questions are worded. We therefore developed the EmojiGrid: a graphical (language-independent) intuitive self-report tool to measure food-related valence and arousal.

In a first experiment participants rated the valence and the arousing quality of 60 food images, using either the EmojiGrid or two independent visual analog scales (VAS). The valence ratings obtained with both tools strongly agree. However, the arousal ratings only agree for pleasant food items, but not for unpleasant ones. Furthermore, the results obtained with the EmojiGrid show the typical universal U-shaped relation between the mean valence and arousal that is commonly observed for a wide range of (visual, auditory, tactile, olfactory) affective stimuli, while the VAS tool yields a positive linear association between valence and arousal. We hypothesized that this disagreement reflects a lack of proper understanding of the arousal concept in the VAS condition.

In a second experiment we attempted to clarify the arousal concept by asking participants to rate the valence and intensity of the taste associated with the perceived food items. After this adjustment the VAS and EmojiGrid yielded similar valence and arousal ratings (both showing the universal U-shaped relation between the valence and arousal). A comparison with the results from the first experiment showed that VAS arousal ratings strongly depended on the actual wording used, while EmojiGrid ratings were not affected by the framing of the associated question. This suggests that the EmojiGrid is largely self-explaining and intuitive.

To test this hypothesis, we performed a third experiment in which participants rated food images using the EmojiGrid without an associated question, and we compared the results to those of the first two experiments. The EmojiGrid ratings obtained in all three experiments closely agree.

We conclude that the EmojiGrid appears to be a valid and intuitive affective self-report tool that does not rely on written instructions and that can efficiently be used to measure food-related emotions.

Introduction

Background

Besides the sensory characteristics of food, food-evoked emotion is a crucial factor in predicting consumer's food preference and therefore in developing new products (Dalenberg et al., 2014; Gutjar et al., 2015b). Hedonic ratings alone do not predict food choice behavior accurately (Wichchukit and O'Mahony, 2010a; 2011). Dalenberg et al. (2014) showed that consumers' emotions add predictive power to a food choice (predicting) model based on hedonic scales, while Gutjar et al. (2015b) found that self-reported food-evoked emotions can predict individual's food choice more accurately than hedonic scores. These studies suggest that assessing emotional responses to foods may reveal product attributes which can be a valuable source of information for product development and marketing that goes beyond traditional sensory and acceptability measurements (Thomson et al., 2010; Jaeger et al., 2018b). Therefore, it seems important to obtain valid and reliable measurements of food-evoked emotions. According to the circumplex model of affect (Russell, 1980b), emotions are characterized by both their valence (pleasantness; the degree of positive or negative affective response to a stimulus) and arousal (the intensity of the affective response to a stimulus; the degree of activation or deactivation). Since hedonic ratings alone do not predict food choice behavior accurately (Wichchukit and O'Mahony, 2010a; 2011), it appears that both valence and intensity play a distinct and critical role in eating-related behavior (Woodward et al., 2017).

Human affective response to food can be assessed objectively by measuring the user's behavioral (e.g., amount consumed, facial expressions) and physiological (e.g., electrodermal activity, heart rate) signals and subjectively using affective self-report tools (e.g., questionnaires, affective lexicons, graphical scales; for a recent review of all different assessment methods see: (Kaneko et al., 2018b)).

Affective self-report questionnaires are the most widely used tools since they are extensively validated and easy to apply. These tools can be divided into two main groups: tools that represent emotions verbally (e.g., through names like "*fear*", adjectives like "*afraid*" or even full sentences: (King and Meiselman, 2010; Spinelli et al., 2014; Nestrud et al., 2016a)) and tools that represent emotions graphically (e.g., through smiling or frowning faces: (Bradley and Lang, 1994; Vastenburg et al., 2011; Laurans and Desmet, 2012a; Broekens and Brinkman, 2013; Huisman et al., 2013; Obaid et al., 2015)).

Verbal tools enable users to report their current affective state by selecting or rating words that best express their feelings. They are the most commonly used techniques to measure emotional responses to food, due to their ease of application, cost-effectiveness and discriminative power (Churchill and Behan, 2010; Dorado et al., 2016). However, they have several shortcomings: *(1)* affect and emotions (especially mixed or complex ones) are difficult to verbalize and the labels used to describe them are inherently ambiguous (Scherer, 2005; Köster and Mojet, 2015a) and *(2)* the 'affective' or 'emotional' lexicon varies across cultures and languages, particularly when it comes to foods (e.g., (Curia et al., 2001; Gutjar et al., 2015b; van Zyl and Meiselman, 2015)). Also, verbal tools are demanding for the user since they require cognitive effort (interpretation) and a significant amount of time to fill them out. This disadvantage increases when they need to be repeatedly applied in the course of an experiment.

Graphical tools allow users to report their feelings efficiently and intuitively by indicating or rating the (part of the) figure that best represents their current affective state. Graphical self-report instruments are appealing for the measurement of affective experiences since they do not require the users to verbalize their emotions. Instead, they rely on the human ability to intuitively and reliably attribute emotional meaning to (simple) graphical elements (Aronoff et al., 1988; Windhager et al., 2008; Larson et al., 2012; Watson et al., 2012), in particular those linked to facial expressions (Tipples et al., 2002; Lundqvist et al., 2004; Weymar et al., 2011). It has therefore been suggested to replace the subjective linguistic increments on rating scales by iconic facial expressions (Kaye et al., 2017). Since graphical self-report tools do not rely on verbal descriptions of emotions, they may also be useful for cross-cultural studies since they eliminate the need for translation and the problems associated therewith (e.g., (Curia et al., 2001; van Zyl and Meiselman, 2015)). Also, they may be more effective to measure and express mixed (complex) emotions that are hard to verbalize (Elder, 2018). Hybrid tools that combine graphical elements with verbal labels to clarify their meaning (e.g., (Cowie et al., 2000)) may be useful for populations with inherent reading problems (e.g., dyslexia).

In the next section, we first give a brief overview of existing affect self-report measurement tools, focusing in particular on pictorial scales, and we discuss their limitations as tools to measure food-related emotional experiences.

Related work

Affective self-report through cartoon characters

The Affect Grid (Russell et al., 1989) is a two-dimensional labeled visual scale to assess affect along the principal dimensions valence and arousal, based on Russell (1980b)’s circumplex model of affect. The horizontal valence scale ranges from “*unpleasant*” (low negative valence) to “*pleasant*” (high positive valence). The vertical arousal scale ranges from “*sleepiness*” (low intensity - no arousal) to “*high arousal*” (high intensity). Four additional labels (“*stress*”, “*excitement*”, “*depression*”, “*relaxation*”) clarify the meaning of the extreme emotions represented by the corners of the grid. Users mark the location on the grid that best corresponds to their affect state after perceiving a given stimulus. Hybrid abstract and pictorial versions of the Affect Grid have been created by labeling its axes either with icons of faces showing different emotional expressions (Schubert, 1999) or with abstract cartoon characters (Swindells et al., 2006; Cai and Lin, 2011). Although the Affect Grid has been applied to measure food elicited emotions (Einöther et al., 2015; den Uijl et al., 2016b), none of these tools has been specifically designed to assess food-related emotions.

Other affective self-report tools use cartoon characters that express specific emotions through facial and bodily expressions. The rationale for their use is twofold. First, people can accurately identify discrete emotions from bodily signals such as facial expressions (Ekman, 1994) and body language (Wallbott, 1998) across cultures (Ekman and Friesen, 1971). Second, visually expressed emotions are hypothesized to more closely resemble intuitively experienced emotions (Dalenberg et al., 2014). Evidence for this hypothesis stems from EEG-experiments showing that emotion processing is faster for facial expressions than for emotional words (Schacht and Sommer, 2009; Frühholz et al., 2011; Rellecke et al., 2011). Although none of the currently available cartoon-based self-assessment tools have been designed to measure food-related emotions, we will first give a brief overview of the existing methods since they are closely related to the new tool that we will present later in Section 2.

The Self-Assessment Mannikin (SAM: (Bradley and Lang, 1994)) is a pictorial assessment technique that enables users to report their momentary feelings of valence, arousal, and dominance by selecting for each factor from a set of humanoid figures showing different intensities the one that best expresses their own feeling. Muñoz et al. (2010) introduced an additional SAM scale to measure food related craving (the desire to consume; see also: (Miccoli et al., 2014)). Although the SAM is widely used and extensively validated, it is generally acknowledged that it has several serious drawbacks. First, users often misunderstand the depicted emotions. Especially children have difficulties understanding the SAM (Yusoff et al., 2013; Hayashi et al., 2016). While the valence dimension of the SAM is quite intuitive (depicted as the figure's facial expression going from a frown to a smile), the dominance dimension (depicted as the size of the figure) is much harder to explain, and the arousal dimension (depicted as an 'explosion' in the stomach area) is often misinterpreted (Broekens and Brinkman, 2013; Betella and Verschure, 2016; Chen et al., 2018b). Second, the method still requires a successive assessment of the stimulus on multiple dimensions separately.

PrEmo (Product Emotion Measurement Instrument) is a non-verbal cross-cultural validated self-report instrument to measure 14 distinct emotions visualized by an animated cartoon character (Desmet et al., 2000a; Laurans and Desmet, 2012a). Users rate to what extent the animated figures express their feelings elicited by a stimulus, using a 5-point scale. Although PrEmo has been applied to measure food elicited emotions (Dalenberg et al., 2014; Gutjar et al., 2015b; den Uijl et al., 2016b; He et al., 2016a; He et al., 2016b), it was not designed for this purpose and most of the displayed emotions (e.g., *pride*, *hope*, *fascination*, *shame*, *fear*, *sadness*) therefore have no evident relation to food experiences. Similar cartoon-based self-report tools representing a limited set of emotions are the Pictorial Mood Reporting Instrument (PMRI: (Vastenburg et al., 2011)), the pictorial ERF (Emotion Rating Figurines: (Obaid et al., 2015)), the LEMtool (Layered Emotion Measurement tool: (Huisman and van Hout, 2008; Huisman et al., 2013)) and Pick-A-Mood (Desmet et al., 2016). The Affective Slider is a digital scale composed of two vertically aligned sliders labeled with stylized facial expressions that represent pleasure and arousal (Betella and Verschure, 2016). Unlike the previous methods, the AffectButton (Broekens and Brinkman, 2013) and EMuJoy (Emotion measurement with Music by using a Joystick: (Nagel et al., 2007)) allow users to continuously adjust the emotional expression of a cartoon character (by moving a mouse controlled cursor). However, these tools require the user to successively explore the entire affective space to find the desired expression each time a response is given, unlike the other graphical tools that provide an instantaneous overview of the affective input space.

Affective self-report through emoji

Emoji are pictographs or ideograms representing emotions, concepts, and ideas. They are widely used in electronic messages and Web pages to supplement or substitute written text (Danesi, 2016). Facial emoji are typically used to change or accentuate the tone or meaning of a message. They can support users to express and transmit their intention more clearly and explicitly in computer mediated communication (dos Reis et al., 2018). Emoji span a broad range of emotions, varying in valence (e.g., smiling face vs. angry face) and arousal (e.g., sleepy face and face with stuck-out tongue and winking eye). Although some facial emoji can be poly-interpretable (Miller et al., 2016; Tigwell and Flatla, 2016) it has been found that emoji with similar facial expressions are typically attributed similar meanings (Moore et al., 2013; Jaeger and Ares, 2017) that are also to a large extent language independent (Kralj Novak et al., 2015). Emoji can elicit the same range of emotional responses as photographs of human faces (Moore et al., 2013). In contrast to photographs of human faces, emoji are not associated with overgeneralization (the misattribution of emotions and traits to neutral human faces that merely bear a subtle structural resemblance to emotional expressions: (Said et al., 2009)), or racial, cultural and sexual biases.

For a study on children's sensitivity to mood in music Giomo (1993) developed a non-verbal response instrument using schematic faces arranged in a semantic differential format along three lines corresponding to each of the three musical mood dimensions defined by Wedin (1972). By marking the most appropriate facial expression children used the tool to report their perceived mood in musical pieces.

Schubert (1999) developed the interactive Two-Dimensional Emotion-Space (2DES) graphic response tool to enable continuous measurement of perceived emotions in music. The 2DES tool consists of a square Affect Grid (with valence along the horizontal and arousal along the vertical axis) with schematic faces (showing only eyes and a mouth) arranged at the corners and the midpoints of the four sides of the grid. No further labels are provided. The human-computer interface records cursor movements within the square. The schematic faces represent the arousal dimension by the size of the mouth and the eye opening, while the valence dimension is represented by the concavity of the mouth. These features are based on the literature on facial expression (Ekman et al., 1971). An extensive evaluation study showed that the instrument was intuitive to use and had a significant reliability and validity (Schubert, 1999). The author suggested that the tool could be applied to measure emotion felt in response to a stimulus rather than emotion expressed by the stimulus (Schubert, 1999).

Russkman (Russell and Ekman**:** (Sánchez et al., 2006)) is an interactive graphic response tool consisting of a set of emoji expressing 28 affective states on three levels of intensity. Russkman is based on Russell (1980b)'s circumplex model of affect and Ekman's facial Action Coding System (FACS: (Ekman and Rosenberg, 2004)) and was developed to convey mood and emotion in instant messaging. The user can select a specific emotion by moving a cursor on top of one of the four icons representing the quadrants of an Affect Grid, which then expands making all icons in this quadrant available for selection.

To make the SAM more accessible to children, Hayashi et al. (2016) replaced the cartoon characters with emoji. Their 5-point 'emoti-SAM' was quickly grasped by children and effectively used as both an online and a paper version.

Swaney-Stueve et al. (2018) developed a 7-point bipolar valence scale labeled with emoji. They compared this scale to a 9-point verbal liking scale in an online experiment in which children reported their affective responses to different pizza flavors and situations. Both scales yielded similar responses distributions with a strong positive linear correlation (R^2>0.99 for both pizza flavors and situations). They concluded that further research was needed to extend their unidimensional emoji scale into a two-dimensional one that also measures arousal.

Emoji based rating tools are increasingly becoming popular tools as self-report instruments (Kaye et al., 2017) to measure for instance user and consumer experience (e.g. www.emojiscore.com). For instance, Moore et al. (2013) developed a 9-point emoji scale to measure users' affective responses to an online training simulation, and Alismail and Zhang (2018) used a 5-point emoji scale to assess user experience with electronic questionnaires. While emoji typically express different degrees of valence and arousal (Moore et al., 2013), previous studies only validated (Aluja et al., 2018) and used (Moore et al., 2013; Alismail and Zhang, 2018) the valence dimension.

While people do not easily name food-related emotions, they appear to use emoji in a spontaneous and intuitive way to communicate food-related emotional experiences (Vidal et al., 2016). Previous studies found that emoji can serve as a direct self-report tool for measuring food related affective feelings (Vidal et al., 2016; Ares and Jaeger, 2017; Gallo et al., 2017; Jaeger et al., 2017c; Jaeger et al., 2018a; Schouteten et al., 2018). However, these previous studies used subsets of the most popular and currently available emoji, most of which show facial expressions that have no clear relation to food experiences. Also, the size of these sets (33 emoji: (Ares and Jaeger, 2017; Jaeger et al., 2017c; Jaeger et al., 2018a; Schouteten et al., 2018), 25-39 emoji: (Jaeger et al., 2017a), and 50 emoji: (Gallo et al., 2017)) is rather overwhelming and comparable to the large number of words typically used in emotional lexicons to measure emotional associations to food and beverages (e.g., (King and Meiselman, 2010; Spinelli et al., 2014; Nestrud et al., 2016a)). These large set sizes make emoji-based rating or selection procedures quite inefficient. Sets of emoji were used in both CATA (check-all-that-apply) (Ares and Jaeger, 2017; Jaeger et al., 2017a; Jaeger et al., 2017c; Schouteten et al., 2018) and RATA (rate-all-that-apply: (Ares and Jaeger, 2017)) questionnaires. In general, these studies found that emoji are capable to discriminate well between hedonically diverse stimuli, while the reproducibility of the emotional profiles was quite high (Jaeger et al., 2017c). Compared with other non-verbal methods that use cartoon figures to represent different emotions (e.g., (Desmet et al., 2012; Laurans and Desmet, 2012a; Huisman et al., 2013)), emoji characters appear to have the advantage of being more familiar to users. It seems that users easily connect emoji to food-elicited emotions, even without any explicit reference to feelings in the wording of the associated question (Ares and Jaeger, 2017). Given that emotions in facial expressions, gestures and body postures are similarly perceived across different cultures (Ekman and Friesen, 1971; Ekman, 1994), cross-cultural differences in the interpretation of emoji could also be smaller than the influences of culture and language on verbal affective self-report tasks (Torrico et al., 2018b). Also, emoji provide a visual display of emotion, making them also beneficial for use with children who may not have the vocabulary to convey all their emotions (Gallo et al., 2017; Schouteten et al., 2018).

For repeated or routine testing in applied settings, selecting emoji from a long list of possible candidates may be a task that is too demanding, and shorter tests are therefore required. The emoji used to measure food-related emotions in previous studies (Ares and Jaeger, 2017; Gallo et al., 2017; Jaeger et al., 2017a; Jaeger et al., 2017c; Schouteten et al., 2018) were not specifically developed for this purpose but were merely selected as the most appropriate ones from the general set of available emoji. As a result, several emoji were obviously out of context and had no relevance for the description food-related affective associations (Jaeger et al., 2017c). Also, the most frequently used emoji are primarily associated with positive emotional experiences reflecting the dominance of positive emotions in food consumption (hedonic asymmetry: (Desmet and Schifferstein, 2008)). Hence, there is a need for a set of emoji that *(1)* specifically relate to food experience and *(2)* that span the entire hedonic continuum from negative to positive emotions.

Current study

In the previous section, we identified a need for an efficient food-specific graphical (language independent) affective self-report method that produces reliable and valid data. We also identified emoji as a promising graphical avenue.

In the rest of this paper, we first introduce and validate the EmojiGrid, which is a new efficient graphical self-report tool that can measure food related affective states along the dimensions of valence and arousal.

Then we will present the results of two comparative evaluation studies in which participants rated valence and arousal of images of food either with the new EmojiGrid scale or with conventional visual analog scales (VAS) scales. Previous research has shown that viewing pictures of food not only activates the visual cortex, but also brain areas that code how food actually tastes (the insula/operculum) and the reward values of tasting it (the orbitofrontal cortex; (Simmons et al., 2005)). Food images are therefore typically considered a viable surrogate for the real thing (e.g., (Foroni et al., 2013; Blechert et al., 2014c; Miccoli et al., 2014; Miccoli et al., 2016)).

Finally, we will present some conclusions and suggestions for future research.

EmojiGrid: Design and validation

Here we propose the EmojiGrid (see Figure 7-1) as a new tool to assess food-related affective associations.

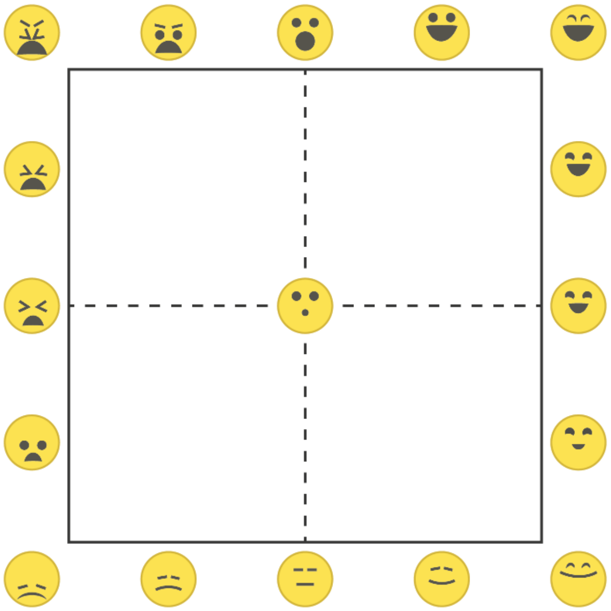


Figure 7-1. The EmojiGrid: an emoji-labeled Affect Grid for the measurement of food-related affective associations.

Design

The EmojiGrid is a Cartesian grid similar to the Affect Grid (Russell et al., 1989), but the verbal labels are replaced with emoji showing food-related facial expressions. Also, additional emoji are inserted between the midpoints and the endpoints of each axis (resulting in five emoji on each side of the grid), and one emoji is placed in the center of the grid, resulting in a total of 17 emoji on the grid. A central neutral expression serves as a baseline or anchor point. The facial expressions vary from disliking (unpleasant) via neutral to liking (pleasant) along the horizontal (valence) axis, and gradually increase in intensity along the vertical (arousal) axis. The facial expressions are defined by the eyebrows, eyes and mouth configuration of the face, and are inspired by the Facial Action Coding System (Ekman and Friesen, 2003). The arousal dimension is represented by the opening of the mouth and the shape of the eyes, while the valence dimension is represented by the concavity of the mouth, the orientation and curvature of the eyebrows, and the vertical position of these features in the face area (representing a slightly downward looking face for lower arousal values and a slightly upward looking face for higher valence values). These facial features represent a minimal set needed to express the range of emotions over the Affect Grid. To avoid potential biases in ratings due to the emotional connotation of colors (Clarke and Costall, 2008; Suk and Irtel, 2010), we adopted a monochromatic (yellow) color scheme in the design of the EmojiGrid. Users place a check mark at the location in the grid that corresponds to the emoji (facial expression) that best represents their affective state (feelings) after perceiving a certain food or beverage.

Previous studies using emoji to measure food evoked emotions typically started with a large set of currently available (extremely heterogeneous) emoji and merely selected those emoji that could somehow be related to food (e.g., (Ares and Jaeger, 2017; Gallo et al., 2017; Jaeger et al., 2017a; Jaeger et al., 2017c; Jaeger et al., 2018a; Schouteten et al., 2018)). This approach typically results in a limited set of emoji with widely different (and not systematically varying) characteristics, that also does not cover the full valence-arousal space. The emoji used to label the EmojiGrid were designed to represent facial expressions corresponding to the emotions represented by the grid points along the outer edges of the Affect Grid that represents the general affective dimensions of valence and arousal. Hence, the iconic facial expressions of the emoji represent emotions that can be induced by any stimulus or event, including food. Thus, the stimuli were not specifically designed to reflect only food induced emotions. The systematic variation in the shape and size of the facial characteristics (eyebrows, eyes and mouth) of the emoji enables users to interpolate facial expressions between the label icons on the edges of the grid.

Validation

We performed three validation studies to assess whether the emoji had indeed the intended and intuitive order across the valence-arousal space. The tasks involved an integral interpretation of the shape and size of the mouth and eyes and the position and shape of the eyebrows. While, as noted before, the stimuli were designed to represent general emotions and not merely reflect food-induced emotions, the facial expression were such that all of them could be related to food induced emotions.

Affective assessment of individual emoji

To validate the EmojiGrid, a convenience sample of 28 Dutch students (18 females, 10 males), aged between 18 and 24, rated each individual emoji label on valence and arousal, using 5-point SAM scales. The emoji were presented in random order.

Pearson's correlation between the SAM valence and arousal ratings and the scale values corresponding to the position of the emoji on the EmojiGrid (i.e., the label indices) was respectively r(15) =0.96 and 0.92, p<0.01, indicating close agreement between both scales. This result agrees with that of Swaney-Stueve et al. (2018), who found that a 7-point valence scale labeled with emoji and a verbal liking scale yielded similar responses distributions with a strong positive linear correlation (R^2>0.99 for both pizza flavors and situations).

In this experiment the emoji were individually presented in random order. In the actual EmojiGrid they are arranged along the edges in order or increasing valence and arousal. We hypothesize that this linear spatial arrangement along the edges will serve to provide a correct impression of the corresponding gradual variation in facial expression (representing either valence or arousal).

Linear ordering emoji of similar valence or arousal

To test the hypothesis that a linear arrangement can enhance the perception of the logical order of the emoji labels on the edges of the EmojiGrid, a convenience sample of 10 Dutch students (5 females, 5 males), aged between 19 and 25, ordered 4 sets of 5 emoji each. The 4 sets represented respectively the emoji labels on the bottom edge (low arousal: Figure 7-2A), top edge (high arousal: Figure 7-2B), left edge (low valence: Figure 7-2C) and right edge (high valence: Figure 7-2D). The participants were asked to order the 2 sets of emoji with similar (low or high) arousal (Figure 7-2A,B) along a line segment in order of increasing valence and the 2 sets of emoji with similar (low or high) valence (Figure 7-2C,D) in order or increasing arousal. The stimuli were presented in PowerPoint slides. The emoji were initially randomly ordered on top of the screen, and the participants could drag them to the lines segment using their mouse.

The two sets of emoji with similar arousal values (Figure 7-2A,B) were both correctly ranked in order of increasing valence by all participants. For the two sets of emoji with similar valence values (Figure 7-2C,D), only the first two emoji (i.e., the ones representing the lowest level of arousal) were ranked in reverse order by 3 (out of 10) participants. This reversal of emoji representing low arousal may be resolved when a valence context is provided (i.e., when the arousal axes are flanked by their corresponding valence axes, as in the EmojiGrid).

Figure 7-2. The four sets of five emoji each used in the ranking experiment, representing the emoji along (A) the bottom edge (lowest arousal), (B) top-edge (highest arousal), (C) left edge (lowest valence), and (D) right edge (highest valence).

Circular ordering of emoji

To test the hypothesis that people are able to correctly order the labels of the EmojiGrid when their full valence-arousal context is provided, we asked participants to arrange the 16 emoji that are used as labels on the edges of the EmojiGrid along a circle (the topological equivalent of the boundary of the EmojiGrid) in a "logical order". The stimuli were presented in PowerPoint slides. The 16 emoji were initially displayed in a random order on the upper part of the screen, and a large circle was shown on the lower part of the screen. The participants could use their mouse to place the emoji anywhere along the circle using their mouse. A convenience sample of 10 Dutch students (5 females, 5 males), aged between 21 and 25, ordered all 16 emoji. Most participants correctly ordered the emoji (i.e., in the same order as they have along the EmojGrid). Two participants reversed the order of the neutral emoji with the highest valence (3rd emoji in Figure 7-2B) and the emoji with the lowest valence and second lowest arousal (2nd emoji in Figure 7-2C).

General methods

Measures

In this study participants rated the valence and arousal of food images using either the new EmojiGrid or labeled VAS scales. The participants responded by positioning a cursor on the appropriate location of the respective scales and clicking with the mouse button. The two dimensions of the EmojiGrid and the two VAS scales (corresponding to the affective dimensions of valence and arousal) were all converted to a range from 0 to 100 points.

Participants

Participants with a Dutch nationality were recruited through postings on social media and direct emailing. The experimental protocol was reviewed and approved by the TNO Ethics Committee (Ethical Application Ref: 2017-011) and was in accordance with the Helsinki Declaration of 1975, as revised in 2013 (World Medical Association, 2013). Participation was voluntary. All participants gave their web-based informed consent instead of written consent. After completing the study the participants were offered to participate in a raffle for vouchers for an online shopping site, with a value of 10 Euros each.

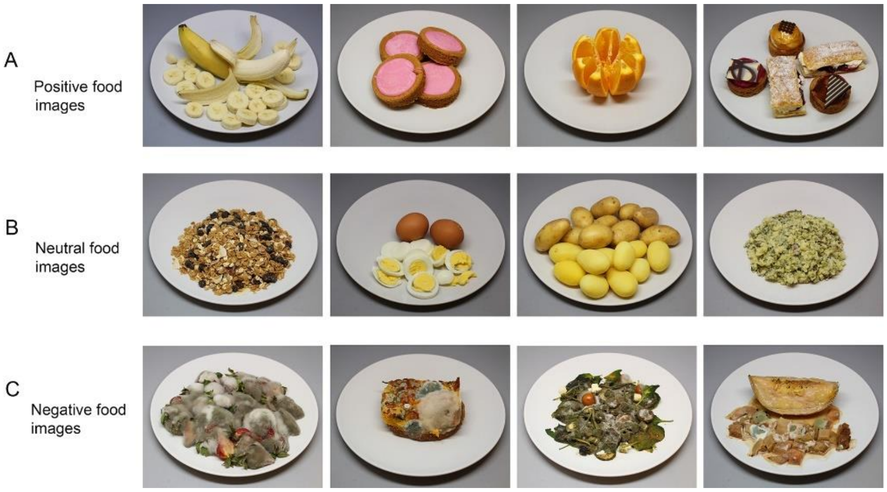


Figure 7-3. Stimulus examples. (A) Positive food images (banana, cookies, orange, sweets), (B) neutral food images (cereals, boiled eggs, boiled potatoes, hotchpot), and (C) negative images of rotten food (strawberries, omelet on bread, Greek salad, melon).

Stimuli

The stimulus set consisted of 60 different food images: 50 images were specifically registered for this study according a standard protocol (see: (Charbonnier et al., 2016); for some examples see: Figure 7-3 and 10 additional images were taken from the FoodCast research image (FRIDa) database ((Foroni et al., 2013); see: Figure 7-4). The 50 images that were registered for this study (see: Figure 7-3) have a resolution of 1037 × 691 pixels and represent natural food (e.g., strawberry, salad), rotten or molded food (e.g., rotten banana, molded salad), raw food (e.g., raw chicken, raw potatoes), processed food (e.g., cakes, fried fish), unfamiliar food (e.g., locusts), and contaminated food (e.g., hotchpot with fake turd). The set of food items was selected such that their perceived valence is likely to be distributed along the entire valence scale (ranging from very low valence for rotten, molded or contaminated food, via neutral for raw onions, boiled eggs or potatoes, to very high valence for fresh fruit, chocolates and pastries). The 10 additional food images from the FRIDa database (Figure 7-4) have a resolution of 530 × 530 pixels and were selected such that their associated valence scores (as reported in their accompanying data file, see: (Foroni et al., 2013)) were approximately evenly distributed over the full range of the valence space covered by this dataset (i.e., ranging from minimal to maximal valence and in between). Five of the images had positive valence (squid/NF_093, ham /TF-087, tacos /TF_141, strawberry /NF_037, tart /TF_093), the other five had negative valence (molded bread /RF_025, sprout /RF_006, Oyster /NF_068, beetroots /NF_015, blue cheese /TF_066). The validated FRIDa images were included as anchor points for verification purposes: in Experiment 1 of the current study, the VAS valence and arousal ratings for these images were obtained following the same procedure as used in the study by Foroni et al. (2013).

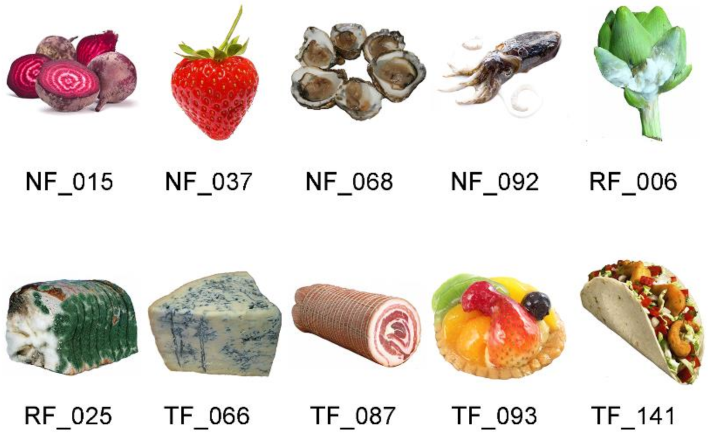


Figure 7-4. The 10 images from the FRIDa database (images reproduced with the permission of the copyright holder: (Foroni et al., 2013)). The labels are the original identifiers of the images in the FRIDa database (NF, natural food; RF, rotten food; TF, transformed food). This set consists of two highly negative (RF) and three highly positive (NF_037, TF_093, and TF_141) images. The remaining five images are distributed over the neutral zone of the valence–arousal continuum.

Procedure

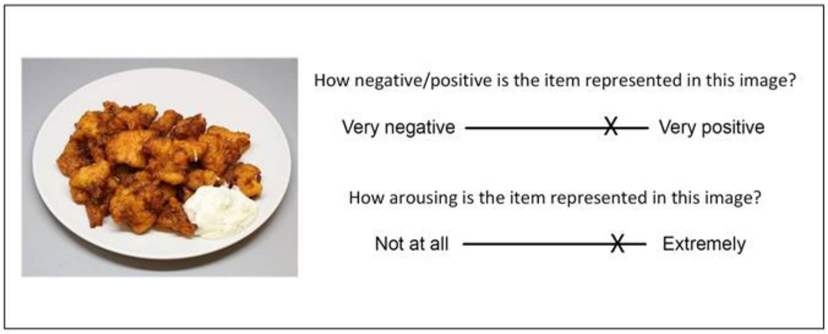
Participants took part in an anonymous online survey. Although internet surveys typically provide less control over the experimental conditions, they typically yield similar results as lab studies (e.g., (Gosling et al., 2004; Woods et al., 2015; Majima et al., 2017)) while they limiting several disadvantages associated with central location studies.

The experiment was programmed in the Java script language, and the survey itself was hosted on a web server. The time stamps of the different events (onset stimulus presentation, response clicks) and the display size and operating system of the participants were logged. This enabled us to check that participants did indeed view the stimuli on larger displays and not on mobile devices with low resolution screens. The resolution of the devices used by the participants in this study varied between 1280x720 and 3440x1440 (the average resolution was 1538x904 pixels across participants, with standard deviations of 330x165 pixels). We could not verify if the browser window was indeed maximized.

The survey commenced by presenting general information about the experiment and thanking participants for their interest. Also, the participants were asked to put their web browser in full-screen mode to maximize the questionnaire resolution and avoid external distractions such as software running in the background. Then the participants were informed that they would see 60 different food images during the experiment and they were instructed to rate their first impression of each image without worrying about calories. It was emphasized that there were no correct or incorrect answers and that it was important to respond seriously. Subsequently, participants electronically signed an informed consent by clicking "*I agree to participate in this study*", affirming that they were at least 18 years old and voluntarily participated in the study. The survey then continued with an assessment of the demographics and the current physical (degree of hunger and thirst, fullness) state of the participants.

Experiment I

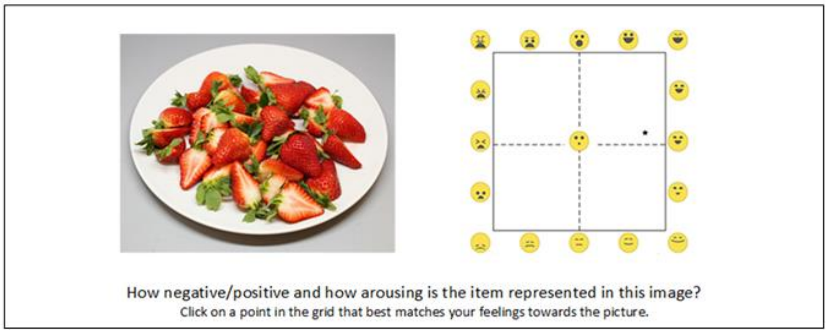
(A) Screenshot of the VAS condition:



How negative/positive is the item represented in this image?

How arousing is the item represented in this image?

(B) Screenshot of the EmojiGrid condition:



How negative/positive and how arousing is the item represented in this image?

Click on a point in the grid that best matches your feelings towards the picture.

Figure 7-5. Screenshot of the VAS (A) and EmojiGrid (B) rating conditions in Experiment I.

Next, the participants were shown either the VAS or the EmojiGrid response tool (depending on the experimental condition to which they were randomly assigned) together with an explanation about how they should use the tool to report their (valence and arousal) ratings for each image.

On each trial the screen displayed the image of a food item on the left side of the screen and the response tool (depending on the experimental condition either a VAS or the EmojiGrid) on the right side of the screen (see: e.g. Figure 7-5). After giving a response by clicking on the (VAS or EmojiGrid) response tool, the next food image appeared, and the response tool was reset (the check mark was removed from the response tool). The presentation duration of each stimulus was not restricted.

The participants first performed two practice trials to familiarize them with the use of the response tool. Immediately after these practice trials, the actual experiment started. The 60 different food images were randomly presented over the course of the experiment. The entire experiment was self-paced and lasted typically about 15 minutes (the mean duration was 14.59 ± 2.32 minutes).

To assess the seriousness of the participation we included a validated seriousness check (Aust et al., 2013). After completing the experiment, participants were asked whether they had answered in a serious manner. The wording of the question was as follows: "*It would be very helpful if you could tell us at this point whether you have taken part seriously, so that we can use your answers for our scientific analysis, or whether you were just clicking through to take a look at the experiment*?" Participants were able to choose one of two answers: "*I have taken part seriously*" or "*I have just clicked through, please throw my data away*." To further motivate the participants, we also included a seriousness question at the start of the experiment: "*It's important for us that you are motivated and answer all questions seriously*". All participants indicated that they took the experiment seriously.

Data analysis

Matlab 2018a (www.mathworks.com) was used to analyze the data and plot the results. For each image and for both self-assessment tools (EmojiGrid and VAS) we computed the mean response across all participants. To get an impression of the agreement between the performance of both response tools we computed the linear correlation between the ratings obtained with both methods for valence and arousal separately. For both tools, we investigated the relation between the valence and arousal ratings by computing least-squares fits of either linear or quadratic functions to the data points.

In this study we checked the occurrence of a random answering behavior by inspecting the consistency of ratings given for stimuli with the highest and lowest overall valence ratings (corresponding respectively to pleasant food items with an overall high positive mean valence rating and unpleasant ones with an overall low negative mean valence rating). We did not observe any outliers, in the sense that there were no participants that gave positive ratings to stimuli with overall mean low valence ratings or the other way around. This suggests that random answering behavior did not occur in this study.

Experiment I: Food valence and arousal measured with EmojiGrid and VAS

This experiment was performed to compare the performance of the EmojiGrid with conventional visual analog scales (VAS) scales. Dutch participants rated the valence and the arousing quality of 60 different food images, using either the EmojiGrid or two independent conventional visual analog scales. The VAS procedure exactly replicated a procedure that was used previously in a similar study in the literature (Foroni et al., 2013) thus enabling to us compare the performance of both methods.

Methods

VAS

In the VAS condition, the participants were asked to rate how each image made them feel by using two scales: one for valence and one for arousal (see Figure 7-5A). The valence scale measured the perceived pleasantness of the displayed product. The question associated with this scale was: "*How negative/positive is the item represented in the image*?" and the extremes of the scale were labeled "*Very negative*" and "*Very positive*". The arousal scale measured the excitement that was experienced while viewing the image. The question associated with this scale was: "*How arousing is the presented image*?" and the extremes of the scale were labeled "*Not at all*" and "*Extremely*". This procedure exactly replicates the one used by Foroni et al. (2013) to assess valence and arousal for the images in the FRIDa database.

The EmojiGrid

In the EmojiGrid condition, participants rated their affective feelings towards each image on the dimensions of valence and arousal by responding to the question "*How negative/positive and how arousing is the item represented in the image*?" using the EmojiGrid, with the additional instruction "*Click on a point in the grid that best matches your feelings towards the picture*" (see: Figure 7-5B).

Participants

The total sample consisted of N = 136 participants, 66 males and 70 females, with a mean age of M = 39.21 (SD = 13.28). Participants were randomly assigned to one of the two experimental conditions.

The sample in the VAS condition consisted of N = 57 participants, 28 males and 29 females with a mean age of M = 42.28 (SD = 13.92).

The sample in the EmojiGrid condition consisbted of N = 79 participants, 38 males and 41 females with a mean age of M = 37 (SD = 12.42).

Results

EmojiGrid versus VAS

For each image and for both self-assessment tools (EmojiGrid and VAS) we computed the mean response across all participants. Figure 7-6 shows the relation between corresponding affective ratings obtained with both tools. This figure suggests an overall linear relation between the valence ratings obtained with both methods, while the relation between the arousal ratings appears to be U-shaped. To get a first impression of the agreement between the performance of both response tools we computed the linear (Pearson) correlation between the ratings obtained with both methods for valence and arousal separately. The valence ratings showed a strong overall positive association between both methods ($r = .97$, $p < .001$). A least-squares linear fit to the data points (blue line in Figure 7-6) confirmed this observation (adjusted R-squared of .92). The arousal ratings showed a moderate and negative association between both methods ($r = -.47$, $p < .001$).

To further quantify the agreement between both methods we also computed intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals, based on a mean-rating, absolute-agreement, 2-way mixed-effects model (see: Table 7-1; (Shrout and Fleiss, 1979; Koo and Li, 2016)). The ICC for valence was 0.958 (with a 95% confidence interval ranging between 0.690 and 0.986) and the ICC for arousal was -0.491 (with a 95% confidence interval ranging between -0.930 and 0.308), indicating that the valence ratings obtained with both methods (EmojiGrid and VAS) are in excellent agreement, while there is no agreement between the arousal ratings.

Table 7-1. Intraclass correlation coefficients for the mean valence and arousal ratings obtained with the VAS and EmojiGrid tools in Experiments I, II, and III.

	Valence	Arousal
VAS_I – VAS_II	0.971 [0.921 – 0.986]	-0.076 [-0.312 – 0.175]
EmojiGrid_I – EmojiGrid_II	0.996 [0.994 – 0.998]	0.954 [0.923 – 0.972]
EmojiGrid_II – EmojiGrid_III	0.998 [0.979 – 0.994]	0.952 [0.911 – 0.973]
EmojiGrid_I – EmojiGrid_III	0.986 [0.975 – 0.992]	0.945 [0.903 – 0.968]

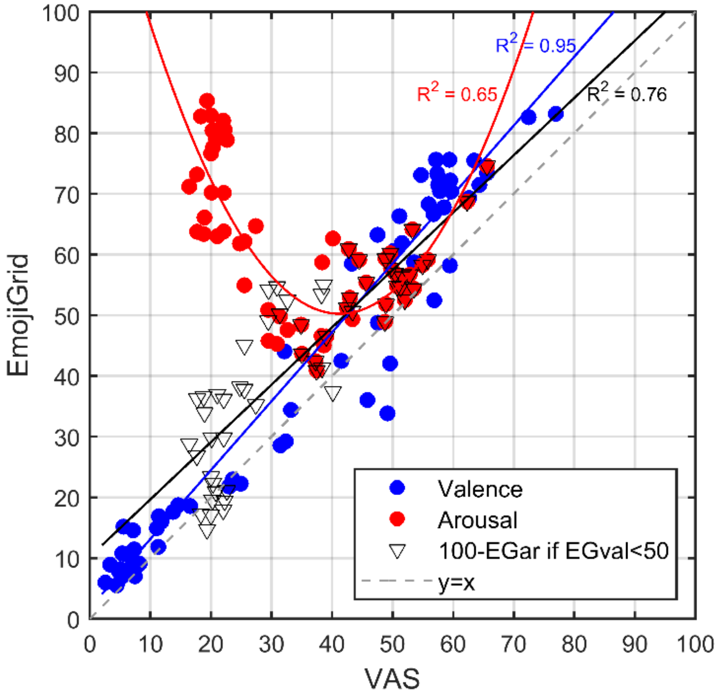


Figure 7-6. Relation between corresponding mean affective ratings obtained with the VAS and EmojiGrid in Experiment I. Blue dots: mean valence ratings. Red dots: mean arousal ratings. Black triangles: mean arousal ratings after inverting the EmojiGrid values for items that were rated as unpleasant (i.e., for which the corresponding valence was lower than 50). The blue and red lines represent linear fits to the valence and arousal ratings. The black line represents a linear fit to the partially inverted EmojiGrid scores. The broken gray line with slope 1 represents full agreement between both methods.

As noted before, Figure 7-6 shows a U- shaped relation between the mean arousal ratings obtained with both rating methods. A least-squares fit showed that the data points are indeed closely approximated by a quadratic function (red line in Figure 7-6; adjusted R-squared = .65). This surprising U-shaped relation suggests that the mean arousal measures resulting from both self-assessment tools may be linearly related if we neglect the polarity (pleasant vs. unpleasant) of the associated valence ratings. To test this hypothesis, we first distributed the arousal measures in two categories based on their corresponding valence values: one category associated with valence values below 50 (images rated as ‘unpleasant’) and one category associated with valence values above 50 (images rated as ‘pleasant’). The arousal ratings obtained with both methods showed a strong negative association ($r = -0.73$, $p < .001$) for images rated as unpleasant and a strong positive association for images rated as pleasant ($r = .78$, $p < .001$). Least-squares linear fits to the left or negative-valenced branch (adjusted R-squared = .52) and the right or positive-valenced branch (adjusted R-squared = .60) of the U-shaped relation between the arousal ratings obtained with both methods showed that the slopes of the negative (-1.1) and positive (0.9) valenced parts of the arousal curve had comparable absolute values. We computed an overall least-squares linear fit to the arousal data points (black line in Figure 7-6; adjusted R-squared = .76) after inverting (subtraction from 100) the arousal values corresponding to images with a valence that was rated below neutral (50). The resulting arousal ratings showed a strong overall positive association between both methods ($r = .87$, $p < .001$).

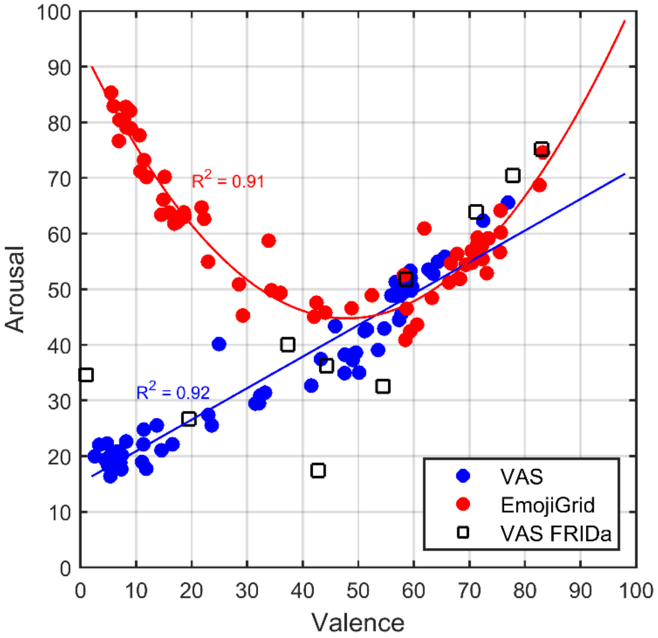


Figure 7-7. Relation between mean valence and arousal ratings for both measurement methods investigated in Experiment I. Blue dots: mean ratings obtained with VAS scales. Red dots: mean ratings obtained with the EmojiGrid. Black squares: mean ratings obtained with VAS scales from the study by Foroni et al. (2013). The blue and red lines represent a linear and quadratic fit to the VAS and EmojiGrid data points, respectively. The adjusted R-squared values represent the agreement between the data and the fits.

Valence versus arousal

The results of Experiment I suggest that the participants interpreted both self-report assessment tools differently for unpleasant images. To follow up on this finding, we plotted the relation between the mean valence and arousal ratings obtained with both self-assessment tools in Figure 7-7. This figure shows the well-known U-shaped relation between valence and arousal for ratings obtained with the EmojiGrid. A least-squares fit showed that a quadratic function closely fits these data points (adjusted R-squared = .91). In contrast, the VAS tool appears to yield a linear relation between valence and arousal ratings. The valence and arousal ratings obtained with the VAS show a strong overall positive association ($r = .96$, $p < .001$). Figure 7-7 also shows the result of a linear least-squares fit to these data points (adjusted R-squared = .92).

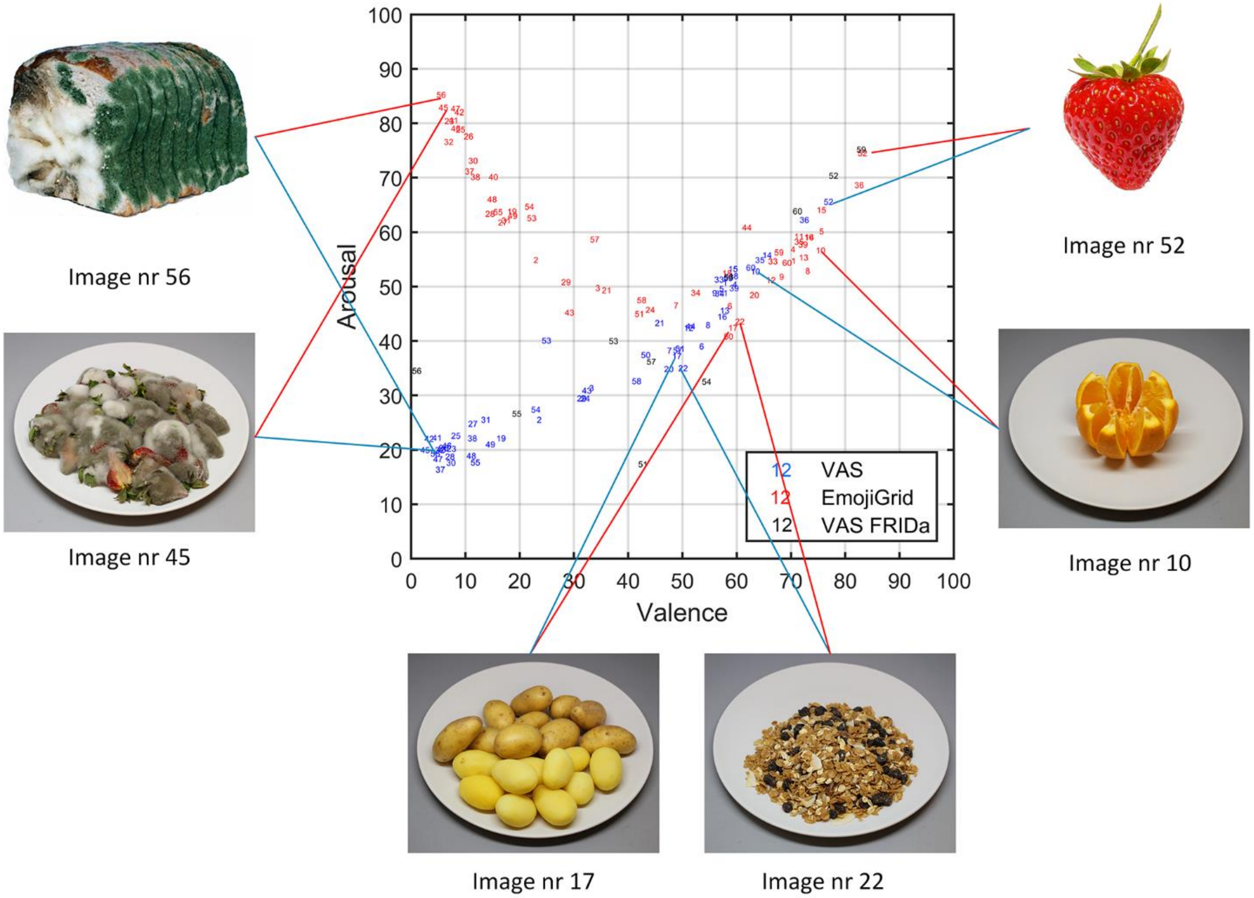


Figure 7-8. As Figure 7-7, where the dot symbols have been replaced by the stimuli indices (i.e., blue numbers represent mean ratings obtained with VAS scales, red numbers correspond to mean ratings obtained with the EmojiGrid, and black numbers represent mean ratings obtained with VAS scales from the study by Foroni et al. (2013)). The VAS and EmojiGrid ratings for appetitive (nr 10 and 52) and neutral (nr 17 and 22) stimuli are similar, while rating for un-appetitive stimuli (nr 45 and 56) are largely different (nr 52 and 56 reproduced with the permission of the copyright holder: (Foroni et al., 2013)).

To check whether the surprising linear relation between valence and arousal for the VAS self-assessment tool is an artifact of the present experimental procedure we compared our results with those of Foroni et al. (2013) who used the same VAS procedure to measure the valence and arousal for the ten FRIDa images included in the present study (represented by the black squares in Figure 7-7). Figure 7-7 shows that the corresponding measurements for these ten images are distributed along the linear least-squares fit to the data points obtained with the VAS tool. To verify this observation, we computed a linear correlation coefficient between the VAS ratings obtained in both studies (i.e., the present study and that of Foroni e.a., 2013) for valence and arousal separately. Both the valence ratings ($r = .87$, $p < .001$) and the arousal ratings ($r = .76$, $p < .001$) obtained with the VAS tool showed a strong overall positive association between both studies. Hence it appears that the results of the VAS tool agree between both studies.

To evaluate the face validity of the valence and arousal ratings we probed which items received extreme (the highest or lowest) and neutral valence and arousal ratings (some examples are shown in Figure 7-8). As expected, both methods yield the highest mean valence ratings for images of fresh fruit, chocolates and pastries, while neutral ratings are obtained for images of raw onions, boiled eggs and potatoes, and the lowest mean valence ratings correspond to images of rotten, molded or contaminated food. However, the arousal ratings obtained with both methods only agree for neutral and positively valenced images, but not for negatively valenced images. Images of rotten, molded or contaminated food yield the highest arousal ratings with the EmojiGrid, but the lowest ratings with the VAS arousal scale.

Discussion

In this experiment, we found a close agreement for the valence ratings provided by the EmojiGrid and VAS tools. The ratings for arousal provided by both methods only agree for pleasant food items, but not for unpleasant ones.

Although the relation between subjective valence and arousal ratings depends both on personality and culture at the idiographic level (i.e., within individuals; (Kuppens et al., 2017)), the shape of this relation is typically characterized by a U-shape at the nomothetic or group level (i.e., both across persons and across a wide range of different stimuli such as sounds, music, paintings, images, movies, words, facial expressions, odors: (Kuppens et al., 2013; Mattek et al., 2017)). The results obtained with the EmojiGrid do indeed reflect this universal U-shaped relation between the valence and arousal ratings, in the sense that the arousal values monotonously increase from the center of the valence scale towards its extremes. Also, arousal values below neutral are scarcely reported, meaning that most food items are typically perceived as stimulating rather than de-activating. However, the VAS tool yields a linear relation between valence and arousal ratings, such that arousal monotonously increases with valence across the entire valence scale. This leads to the surprising result that food items with the lowest perceived valance are rated as least arousing. Our current finding agrees with previous (hitherto unexplained) findings that normative affective picture ratings obtained with labelled continuous VAS slider scales show a linear relation between valence and arousal (Foroni et al., 2013; Blechert et al., 2014c), whereas ratings obtained with a SAM scale show a U-shaped relation (Marchewka et al., 2014; Riegel et al., 2017).

In this experiment, participants rated their feelings towards displayed food items on the affective dimensions of valence and arousal using two different tools (the VAS and EmojiGrid). For both tools, the meaning of the valence dimension was probably evident and directly related to the perceived pleasantness of the displayed products: the VAS tool clearly asked participants to rate the perceived positivity/negativity, and the facial expressions of the emoji in the EmojiGrid clearly displayed (dis-)pleasure. However, the understanding of the meaning of the arousal scale may have differed between both experimental conditions. Since the intensity of the facial expressions clearly increases in the upward direction along the vertical (arousal) axis of the EmojiGrid for each position along its horizontal (valence) axis, the participants probably correctly interpreted this dimension as the intensity of the associated stimulus valence. However, in the VAS condition, the participants may not have understood the meaning of the arousal concept in a food-related context. It has indeed previously been reported that people experience problems understanding the concept of arousal in the context of affective appraisal of food images (Ribeiro et al., 2005). In the current study, such a lack of understanding may have stimulated participants to copy their valence rating when responding their arousal rating. The spatial layout of the VAS may also have promoted such an answering bias: the two scales were presented one above the other, with the valence scale on top. Thus, the arousal response was always given after the valence rating and required a downward mouse movement. If participants were not sure about the meaning of ‘arousal’ this layout made it even more attractive to minimize mouse movements and just click at the same horizontal position on both scales.

Experiment II: Food taste and intensity measured with EmojiGrid and VAS

In Experiment I we found that the EmojiGrid and VAS tools closely agree for the ratings of subjective valence. However, the ratings of subjective arousal only agreed for positively valenced (pleasant) images, but not for negatively valenced (unpleasant) ones. We hypothesized that this disagreement might reflect a lack of understanding of the meaning of the arousal concept.

The instructions used in Experiment I were image-focused and not internal state-focused: we asked for the affective qualities of the food items (how negative/positive and how arousing they were) but not for their immediate impact on the core affective state of the participants. It is however known that the affective qualities of stimuli are differently processed depending on whether they are relevant to the self or not ((Schmitz and Johnson, 2007; Walla et al., 2013); see also (Scherer, 2005)).The appraisal of stimuli with self-relevance stimulates participants to assess their core affective state after engaging in a situated conceptualization based on stored representations of prior experiences (i.e., imagining an experience based on memories and knowledge; (Lindquist et al., 2012)). As a result self-relevance typically enhances the intensity and variation of subjective affective responses (Walla et al., 2013). In this experiment we attempt to enhance the self-relevance of the task by asking participants to rate (i.e., imagine) the expected taste of the stimuli. We hypothesize that this will enhance the perceived arousal for negative valenced stimuli.

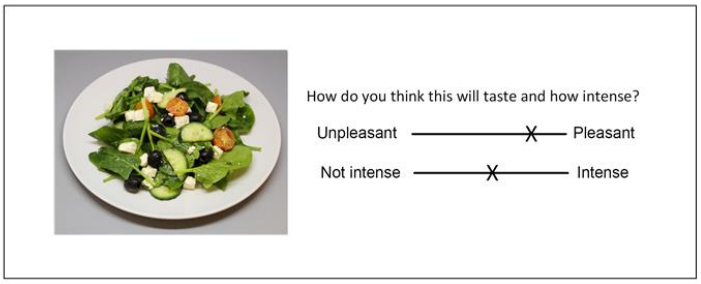
In the context of chemosensory (odor, taste, flavor) stimuli, valence typically measures the hedonic dimension (pleasantness), while self-rated arousal (the subjectively perceived internal state of activation or deactivation engendered by a stimulus) is strongly correlated with the perceived stimulus intensity (Bensafi et al., 2002; Winston et al., 2005; Wang et al., 2016a). Perceived pleasantness and intensity are mediated by different brain mechanisms (Anderson et al., 2003; Small et al., 2003; Cunningham et al., 2004; Grabenhorst and Rolls, 2008). The orbitofrontal cortex evaluates the pleasantness of taste stimuli while the insular taste cortex processes the intensity and identity of the stimulus (Grabenhorst and Rolls, 2008; Grabenhorst et al., 2008). As a result, high level cognitive inputs (de Araujo et al., 2005; Grabenhorst et al., 2008) and selective attention to the affective or physical properties of a stimulus (Veldhuizen et al., 2007; Grabenhorst and Rolls, 2008) differentially modulate the subjectively perceived pleasantness and intensity of taste stimuli.

The way instructions are formulated may therefore well affect the resulting ratings. Studies on the chemical senses typically adopt perceived intensity as a proxy for arousal (e.g. (Small et al., 2003; Cunningham et al., 2004; de Araujo et al., 2005; Grabenhorst and Rolls, 2008; Grabenhorst et al., 2008; Rolls and Grabenhorst, 2008; He et al., 2016a)). In this experiment we will follow this convention and we ask participants to rate not only the perceived valence (pleasantness) but also the perceived intensity of the expected taste of the stimuli. Note that this contrasts with the prevailing definition in emotion theory, where intensity is defined as the length (Euclidian norm) of the vector (with components along the two orthogonal circumplex dimensions valence and arousal) representing a given emotional state (Reisenzein, 1994).

In this experiment, we attempted to simultaneously clarify the meaning of the arousal concept and enhance the arousal response to negatively valenced stimuli in the VAS condition by asking participants to report the expected (imagined) valence and intensity of the taste associated with the perceived food item. We used exactly the same instructions in both (VAS and EmojiGrid) conditions. We hypothesized that this adjusted procedure would lead to a closer agreement between the subjective valence and arousal ratings obtained with both methods.

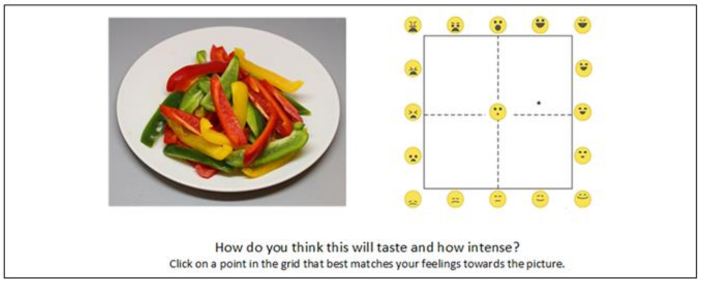
Experiment II

(A) Screenshot of the VAS condition:



How do you think this will taste and how intense?

(B) Screenshot of the EmojiGrid condition:



How do you think this will taste and how intense?
Click on a point in the grid that best matches your feelings towards the picture.

Figure 7-9. Screenshot of the VAS (A) and EmojiGrid (B) rating conditions in Experiment II.

Materials and Methods

VAS

In the VAS condition, the participants were asked to rate how each image made them feel by responding to the question *"How do you think this will taste and how intense?"* using two scales (see: Figure 7-9A): one for valence and the other for arousal. The extremes of the valence scale were labeled “*Unpleasant*” and “*Pleasant*” and the extremes of the arousal scale were labeled “*Not intense*” and “*Intense*”.

The EmojiGrid

In the EmojiGrid condition (see: Figure 7-9B), participants rated their affective feelings towards each image on the dimensions of valence and arousal by responding to the same question “*How do you think this will taste and how intense?”* using the EmojiGrid, with the additional instruction “*Click on a point in the grid that best matches your feelings towards the picture*”.

Participants

The total sample consisted of N = 117 participants, 45 males, and 72 females, with a mean age of M = 32.73 (SD = 15.72).

The sample in the VAS condition consisted of N = 58 participants, 29 males and 29 females with a mean age of M = 32.76 (SD = 15.06).

The sample in the EmojiGrid condition consisted of N = 59 participants, 16 males and 43 females with a mean age of M = 32.69 (SD = 16.47).

Results

For each stimulus and for both self-assessment tools (EmojiGrid and VAS) we computed the mean response across all participants.

Effect of wording (Experiment I versus II)

To investigate whether the wording of the questions associated with both self-report tools affected the ratings, we separately computed the linear correlation between the valence and arousal ratings obtained in Experiments I and II. For the EmojiGrid, both mean valence (r = .98, p < .001) and arousal (r = .95, p < .001) ratings showed a strong overall positive association between both experiments, indicating that the wording of the questions had little or no effect on the subjective ratings obtained with this tool. For the VAS, the mean valence (r = .98, p < .001) ratings also showed a strong overall positive association between both experiments, but the arousal ratings showed no agreement (r = -0.1, p = 0.4).

To further quantify the effect of the wording used for the associated questions on both rating tools we also computed intraclass correlation coefficient (ICC) estimates for the mean valence and arousal ratings obtained with the VAS and EmojiGrid used in Experiment I and II. The results (listed in Table 7-1) show that the valence and arousal ratings obtained with the EmojiGrid are independent of the actual wording used for the associated questions. The same holds for the valence ratings obtained with the VAS. However, there is no agreement between the arousal ratings between both experiments. Hence, it appears that the actual wording used its associated question strongly affects the outcome of the VAS arousal scale.

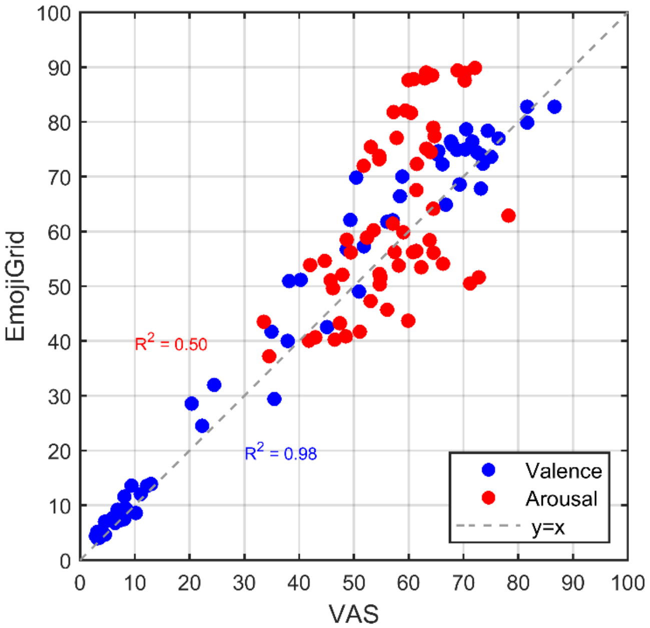


Figure 7-10. Relation between corresponding mean VAS and EmojiGrid ratings in Experiment II. Blue dots: mean valence ratings. Red dots: mean arousal ratings. The broken gray line with slope 1 represents full agreement between both methods. The adjusted R-squared values represent the agreement between the data and a linear fit with slope 1.

Emojigrid versus VAS

Figure 7-10 shows the relation between the mean valence and arousal scores obtained with both the VAS and the Emojigrid in Experiment II. This figure clearly shows an overall linear relation between the ratings obtained with both methods, both for valence and now also for arousal. To illustrate this finding we computed a linear fit with slope 1, which yielded adjusted R-squared values of respectively .98 and .50 (see: Figure 7-10).

To further quantify the agreement between both methods we also computed intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals. The results (listed in Table 7-1) show that the valence ratings obtained with both (EmojiGrid and VAS) methods are in excellent agreement, while there is good agreement between the arousal ratings.

Valence versus arousal

Figure 7-11 shows the relation between the mean valence and arousal ratings obtained in Experiment II with both self-assessment tools. This time we find the well-known U-shaped relation between valence and arousal measurements (Kuppens et al., 2013; Mattek et al., 2017), both for the VAS and the EmojiGrid tools. Least-squares fits to the valence and arousal ratings obtained with both methods show a strong quadratic relation for the EmojiGrid (adjusted R-squared = .89) and a significant quadratic relation for the adapted VAS tool (adjusted R-squared = .37). For comparison, we also plotted the VAS results for the ten FRIDa images from the study of Foroni et al. (2013) in Figure 7-11. It is evident from these results that the adjusted VAS tool used in this experiment attributes significantly higher arousal values to images that are rated as unpleasant compared to the VAS tool used in Experiment I and in the study of Foroni et al. (2013).

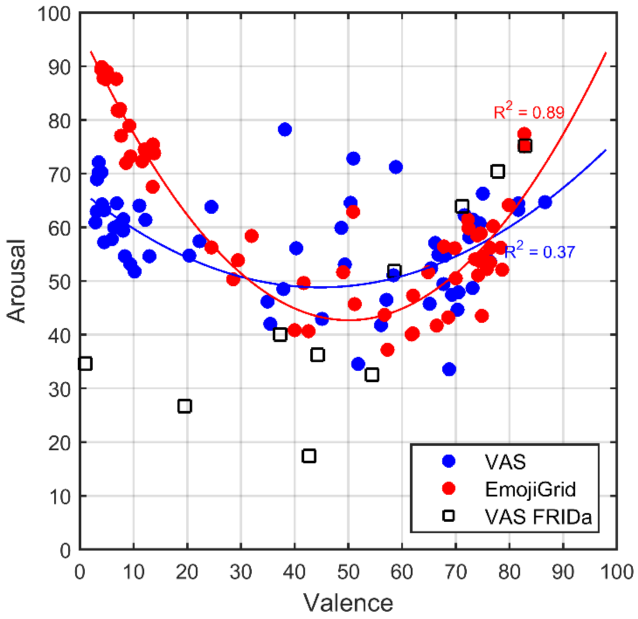


Figure 7-11. Relation between mean valence and arousal ratings for both measurement methods investigated in Experiment II. Blue dots: ratings obtained with VAS scales. Red dots: ratings obtained with the EmojiGrid. Black squares: ratings obtained with VAS scales from the study by Foroni et al. (2013). The blue and red lines represent quadratic fits to the VAS and EmojiGrid data points, respectively. The adjusted R-squared values represent the agreement between the data and the quadratic fits.

To evaluate the face validity of the valence and arousal ratings we again probed which items received extreme (the highest or lowest) and neutral valence and arousal ratings (some examples are shown in Figure 7-12). This time, both methods yield expected and similar results for both valence and arousal: the highest mean ratings are obtained for images of fresh fruit, chocolates and pastries, while neutral ratings are obtained for images of raw onions, boiled eggs and potatoes, and the lowest ratings correspond to images of rotten, molded or contaminated food.

Discussion

In this experiment, we attempted to clarify the meaning of the arousal concept by asking participants to rate the expected intensity of the taste associated with the perceived food item. In addition, we tried to enhance the self-relevance of the task by asking participants to rate (i.e., imagine) the expected taste of the stimuli. We hypothesized that these measures would serve to enhance the perceived arousal for negative valenced stimuli. We found that these procedural adjustments *(1)* indeed raised the mean perceive arousal levels of negatively valenced stimuli and *(2)* resulted in a closer agreement between the subjective valence and arousal ratings obtained with both the VAS and the EmojiGrid tools: both tools now yielded a U-shaped overall relation between the mean valence and arousal curves. Ratings obtained with the VAS arousal scale strongly depended on the actual wording used for its associated question. In contrast, the ratings obtained with the EmojiGrid were not affected by the framing of the associated question. This suggests that the EmojiGrid may be largely self-explaining and intuitive.

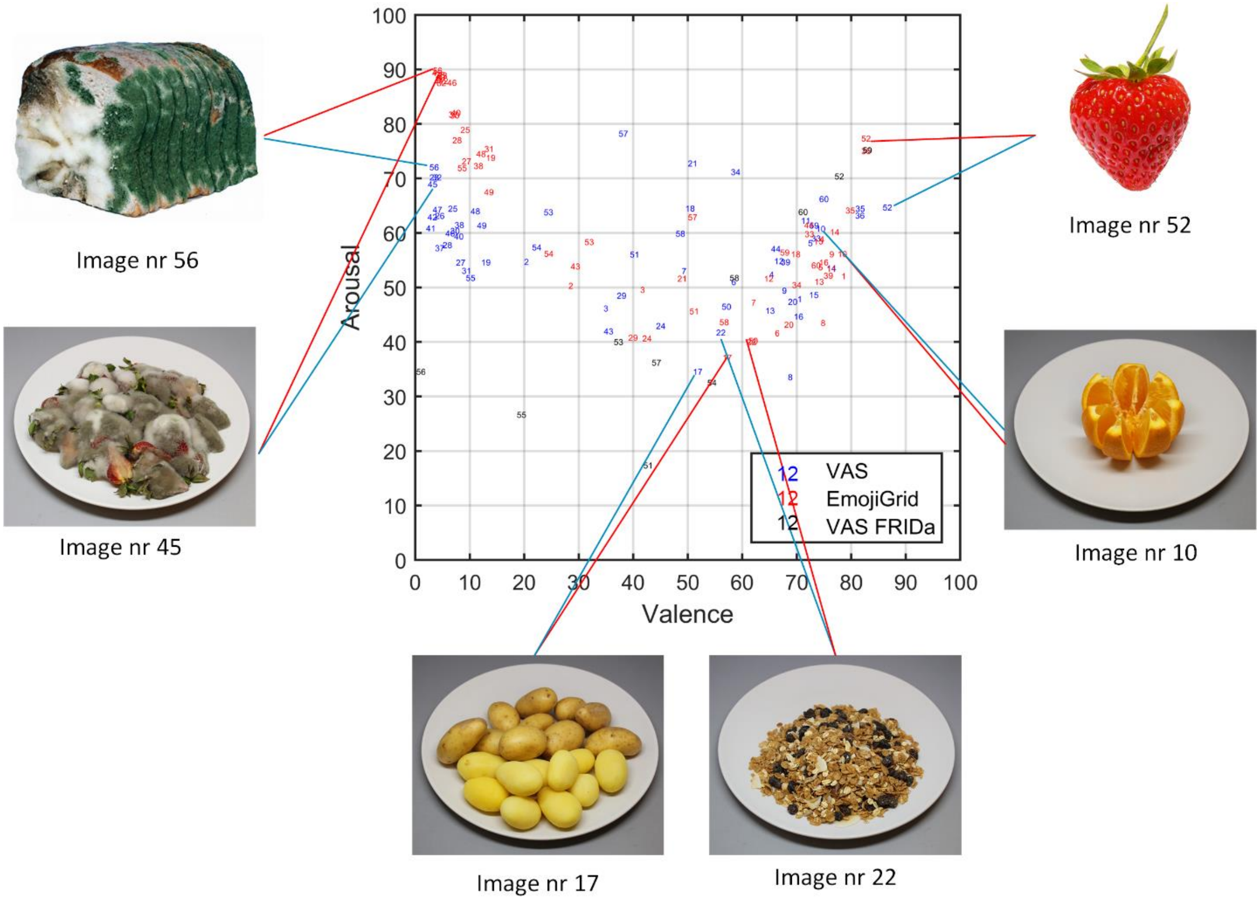


Figure 7-12. As Figure 7-11, where the dot symbols have been replaced by the stimuli indices (i.e., blue numbers represent ratings obtained with VAS scales, red numbers correspond to ratings obtained with the EmojiGrid, and black numbers represent ratings obtained with VAS scales from the study by Foroni et al. (2013). The VAS and EmojiGrid ratings are now similar for appetitive (nr 10 and 52), neutral (nr 17 and 22), and un-appetitive (nr 45 and 56) stimuli (nr 52 and 56 are reproduced with the permission of the copyright holder: (Foroni et al., 2013)).

Experiment III: Affective food response measured with EmojiGrid

To test the hypothesis that the EmojiGrid may be largely self-explaining and intuitive participants rated food pictures online using the EmojiGrid after minimal practice and without any further instructions, and we compared the results with those obtained in Experiment I (where participants were explicitly asked to rate the perceived valence and arousal of the food items) and Experiment II (where participants were explicitly instructed to rate the imagined taste valence and intensity of the food pictures).

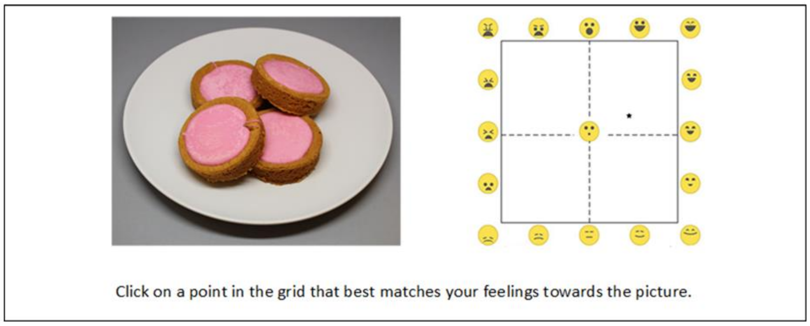
Methods

Procedure

In this experiment participants simply responded their affective feelings towards each image by clicking on an EmojiGrid that was presented without any further verbal instructions (Figure 7-13A). Before starting the actual experiment, they first performed two practice trials to familiarize them with the use of the EmojiGrid. The EmojiGrid in the practice trials was accompanied by the verbal instruction: "*Click on a point in the grid that best matches your feelings towards the picture*" (Figure 7-13A). This instruction was not shown in the actual experiment (Figure 7-13B).

Experiment III

(A) Screenshot of practice trial:



Click on a point in the grid that best matches your feelings towards the picture.

(B) Screenshot of actual trial:

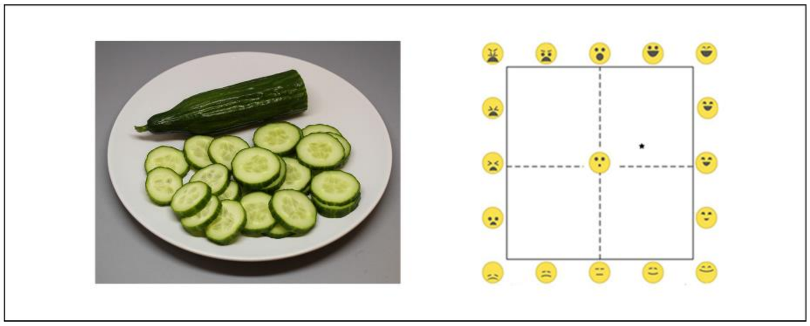


Figure 7-13. Screenshot of a practice trial (A) and an actual trial (B) in Experiment III.

Participants

The total sample consisted of N = 62 participants, 38 males, and 24 females, with a mean age of M = 27.16 (SD = 14.32).

Results

Figure 7-14 shows the relation between the mean valence and arousal ratings obtained with the EmojiGrid in this experiment, together with the previous results from Experiments I and II. This figure shows that the results for all three conditions closely agree. To quantify this agreement, we computed intraclass correlation coefficient (ICC) estimates for the mean valence and arousal ratings obtained in the three different experimental conditions. The results (listed in Table 7-1) show that the valence and arousal ratings provided by the EmojiGrid are in excellent agreement between the different experimental conditions, independent of the presence or the wording of the instructions. This result agrees with the observation of Ares and Jaeger (2017) who found that question wording had little or no effect on affective food evaluation with emoji-based questionnaires.

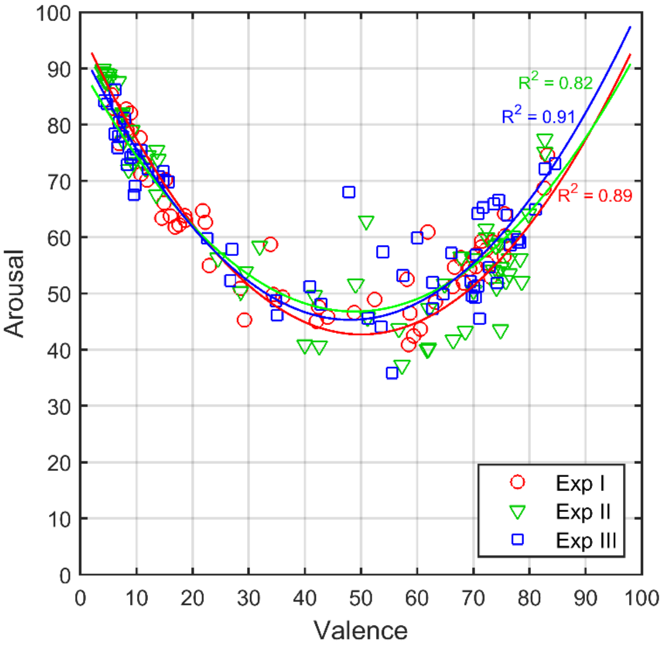


Figure 7-14. Relation between mean valence and arousal ratings obtained with the EmojiGrid in Experiments I, II, and III. The curved lines represent quadratic fits to the data points. The adjusted R-squared values represent the agreement between the data and the quadratic fits.

Discussion

The results from this experiment confirm our hypothesis that the EmojiGrid is largely self-explaining. Given the excellent agreement between the results of the first two experiments (EmojiGrid with instructions) and the third experiment (EmojiGrid without any explanation) it appears that users correctly interpret the valence and arousal dimensions, even without any explanation. It appears that the EmojiGrid is an intuitive instrument that requires no additional associated instructions (referring to either valence and arousal or taste and intensity) apart from the initial instructions to click on a point in the grid that corresponds to the user's current feeling.

General discussion and conclusions

How we feel about food determines to a large extent what, when and how much we eat. Food evaluation studies therefore typically measure the principal affective dimensions of valence and arousal (e.g., (Gil et al., 2009; Esteves et al., 2010; Swan et al., 2013; Miccoli et al., 2014; Piqueras-Fiszman et al., 2014; Hebert et al., 2015; den Uijl et al., 2016a; Wang et al., 2016a; Woodward et al., 2017)). Measures of food-evoked emotions are therefore an essential and valuable source of information for product development and marketing. Hence there is a need for an efficient food-specific self-report tools that produce reliable and valid data. In this paper, we introduced the EmojiGrid as a promising new efficient graphical self-report tool to measure food related affective associations. The EmojiGrid is a Cartesian grid that is labeled with emoji showing food-related facial expressions. Users can report their subjective ratings of valence and arousal by marking the location on the grid that corresponds to the emoji (facial expression) that best represents their affective state after perceiving a given food or beverage. The tool is both intuitive (the facial expressions speak for themselves and don’t need additional labels) and efficient (the two affective dimensions are measured with a single response).

In this study we performed three experiments to validate the EmojiGrid as a self-report tool for measuring food evoked affective feelings. In summary, the aims and key findings of these three experiments are as follows. In two comparative evaluation studies, we first compared the performance of the EmojiGrid with conventional visual analog scales (VAS) scales. The results of the first experiment showed that the valence ratings provided by the EmojiGrid closely agree with those provided by a standard VAS tool, whereas the arousal ratings provided by both methods only agreed for pleasant food items but not for unpleasant ones. Unlike the EmojiGrid, the VAS ratings did not show the universal U-shaped relation between the mean valence and arousal ratings at the group level that is typically reported in the literature. We hypothesized that this disagreement probably resulted from a lack of the participants’ understanding of the arousal concept. In a follow-up experiment, we attempted to clarify the meaning of the arousal concept and to enhance its self-relevance by asking for the expected intensity of the taste associated with the perceived food item. After this adjustment, the valence and arousal ratings obtained with both tools (VAS and EmojiGrid) agreed more closely and both showed the universal U-shaped relation between the valence and arousal. In a final (third) experiment we established that the EmojiGrid yields valence and arousal ratings that do not depend on the actual wording or presence of further instructions. This result contrasts with the finding that ratings obtained with VAS arousal scales strongly depend on the exact formulation of the associated question.

Cross-cultural studies on food-related emotions are becoming increasingly important as a result of the globalization of food products (Meiselman, 2013a). However, verbal self-assessment tools typically pose difficulties for cross-cultural research since emotion words are often not directly equivalent in different languages (Wierzbicka, 1999). In addition, consumers from different cultures tend to use emotions terms differently (van Zyl and Meiselman, 2015; 2016b). The non-verbal and intuitive EmojiGrid may be a valuable tool for cross-cultural studies since it is independent of language and requires only minimal initial instructions ("*Click on the grid*"), exploiting the fact that facial expressions of emotions (e.g., joy, disgust) are largely universal.

Jaeger et al. (2018c) recently found that the use and interpretation of emoji is not influenced by age or frequency of emoji use, suggesting that the EmojiGrid may be a useful tool for users of all ages.

Limitations of this study

The emoji used for the EmojiGrid in this study all had the same size, shape and color. Only their facial (mouth and eyes) features were varied systematically and in a straightforward (simple) way to create various general emotional expressions. It may be possible to design emoji (possibly more elaborated and created by cartoon artists) that are more food-related. Future studies should investigate the effects of graphical emoji properties like size, shape and color on the interpretation of their facial expressions and ultimately on the resulting affective ratings.

The neutral emoji label in the middle of the grid may have had a repulsive effect on the observer response (people may hesitate to click on a face), thus causing a greater variation in the data for (near) neutral stimuli. Future experiments could investigate whether a neutral midpoint is essential.

Future research

This study suggests that the EmojiGrid can indeed capture the affective dimensions of an emotional response to food. Whether such a measure does indeed enable a better prediction of food choice than a unidimensional hedonic rating should be the topic of future studies.

An obvious extension to the present research will be to use the EmojiGrid in food evaluation studies in cross-cultural studies. This involves the investigation of cultural influences on the interpretation of emoji meaning. Emoji may in principle elicit a more intuitive and affective response, which may be particularly useful when testing Asian populations who may be culturally biased to avoid negative scale anchors ((Chen et al., 1995; Lee et al., 2002), see also (Jaeger et al., 2018a))

Children are an important consumer group with special needs. Currently there is no tool for the assessment of children's emotional associations with food (Gallo et al., 2017). Emoji provide a visual display of emotion, making them in principle also a useful tool for populations such as children who do not have the vocabulary to express their emotions. Initial studies have indeed shown that children are quite capable and like to use emoji to characterize their emotions in relation to food (Gallo et al., 2017).

The EmojiGrid may also be a useful tool to evaluate other affective stimuli such as photographs, paintings, music, smells and tactile signals etc. In consumer research, the EmojiGrid can also be used to assess the emotional response to for instance oral care products ((Chen et al., 2018b)), fragrances (Churchill and Behan, 2010), fabrics (Wu et al., 2011), affective ambiences or servicescapes (Kuijsters et al., 2015), etc.

Similar to the AffectButton (Broekens and Brinkman, 2013) and EMuJoy (Nagel et al., 2007), the EmojiGrid may enable users to continuously report perceived affect in human-computer interaction studies by moving a mouse controlled cursor over the support of the grid. While these existing tools require the user to successively explore the entire affective space to find the desired expression each time a response is given, the EmojiGrid provides an instantaneous overview of the affective input space. This feature may be useful for the affective annotation of multimedia (Runge et al., 2016) or personalized affective video retrieval (Xu et al., 2008; Lopatovska and Arapakis, 2011), for real-time affective evaluation of entertainment (Fleureau et al., 2012) or as an affective input tool for serious gaming applications (Anolli et al., 2010).

Acknowledgements

The authors thank Jef van Schendel for designing the emoji labels for the EmojiGrid.

Supplemental material

The EmojiGrid, the 60 food images used as stimuli in the experiments, full documentation of the experimental procedures, and an Excel file with the results are all available from the Figshare Repository (https://figshare.com/articles/_/6463151) with DOI: 10.6084/m9.figshare.6463151.

Chapter 8

EmojiGrid: A 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food

This chapter is published as:
Kaneko, D., Toet, A., Ushiama, S., Brouwer, A.M., Kallen, V., and Van Erp, J.B.F. (2019). EmojiGrid: a 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food. Food research international 115, 541-551.

Abstract

Because of the globalization of world food markets there is a growing need for valid and language independent self-assessment tools to measure food-related emotions. We recently introduced the EmojiGrid as a language-independent, graphical affective self-report tool. The EmojiGrid is a Cartesian grid that is labeled with facial icons (emoji) expressing different degrees of valence and arousal. Users can report their subjective ratings of valence and arousal by marking the location on the area of the grid that corresponds to the emoji that best represent their affective state when perceiving a given food or beverage. In a previous study we found that the EmojiGrid is robust, self-explaining and intuitive: valence and arousal ratings were independent of framing and verbal instructions. This suggests that the EmojiGrid may be a valuable tool for cross-cultural studies.

To test this hypothesis, we performed an online experiment in which respondents from Germany (GE), Japan (JP), the Netherlands (NL) and the United Kingdom (UK) rated valence and arousal for 60 different food images (covering a large part of the affective space) using the EmojiGrid. The results show that the nomothetic relation between valence and arousal has the well-known U-shape for all groups. The European groups (GE, NL and UK) closely agree in their overall rating behavior. Compared to the European groups, the Japanese group systematically gave lower mean arousal ratings to low valenced images and lower mean valence ratings to high valenced images. These results agree with known cultural response characteristics.

We conclude that the EmojiGrid is potentially a valid and language-independent affective self-report tool for cross-cultural research on food-related emotions. It reliably reproduces the familiar nomothetic U-shaped relation between valence and arousal across cultures, with shape variations reflecting established cultural characteristics.

Introduction

The need for a language-independent emotion assessment tool

Our feelings and preferences towards food are significantly determined by our cultural background (Rozin, 1988). How we feel about food determines to a large extent what, when and how much we eat. Food-evoked emotions are therefore a crucial factor in predicting consumer's food preference and in developing and marketing new products (Thomson et al., 2010; Dalenberg et al., 2014; Gutjar et al., 2015b). Because of the globalization of world food markets, cross-cultural research on food-related emotion is becoming increasingly relevant (Meiselman, 2013a). Internet-based surveys appear a promising approach since they provide almost instantaneous access to large consumer groups all over the world (Slater and Yani-de-Soriano, 2010). However, to afford online emotional research, there is a need for validated and efficient culturally independent emotion assessment tools.

Self-response scales are currently by far the most popular method to assess food-evoked emotions (Kaneko et al., 2018b). However, a direct comparison of affective food ratings across cultures is typically compromised by the differential influence and interpretation of verbal scale anchors and by differences in response style (Ares, 2018). It is also stated that the 'emotional lexicon' varies across cultures and languages, particularly when it comes to foods (Gutjar et al., 2015b). Therefore, it is difficult to assure linguistic equivalence of response scales across cultures. On the other hand, Emoji appear a promising alternative for verbal scale labels since they are language independent and intuitive. Emoji are pictographs representing emotions, concepts, and ideas, and are widely used in electronic messages and Web pages to supplement or substitute written text (Danesi, 2016). While verbal labels trigger analytical and rational responses, emoji afford a more intuitive and affective response. This suggests that emoji-based response scales also hold particular promise for Asian consumers (Jaeger et al., 2018a), who are more characterized by holistic thinking (Nisbett et al., 2001). Typical response styles are the extreme response style (ERS) and the middle response style (MRS); that is, the tendency to use either the extremes or the middle parts of the response scales (Harzing, 2006). European and American respondents typically show an ERS, whereas Asian respondents tend to be reluctant to express negative opinions and to provide extreme responses and as a consequence seem to typically show an MRS (Chen et al., 1995; Lee et al., 2002; Lottridge et al., 2012). For instance, Japanese participants tend to use a smaller range of hedonic scales than for instance American participants and tend to avoid the extremes of the scale (Yeh et al., 1998; Yao et al., 2003). While Asians tend to avoid negative and extreme responses when using verbal self-report tools (Chen et al., 1995; Lee et al., 2002; Yao et al., 2003), they appear to freely associate negative feelings to products when using emoji-based self-report tools (Ares and Jaeger, 2017; Jaeger et al., 2017a; Jaeger et al., 2017c; Jaeger et al., 2018a). Emoji-based response tools may therefore inhibit cross-cultural response biases evoked by more verbal assessment methods.

Although there is a wide range of experiential, physiological and behavioral measures of emotion, there is no ‘gold standard’ (Mauss and Robinson, 2009)). The widely used self-response scales can be divided into two main groups: tools that represent emotions verbally (King and Meiselman, 2010; Spinelli et al., 2014; Nestrud et al., 2016a) and tools that represent emotions graphically (Bradley and Lang, 1994; Vastenburg et al., 2011; Laurans and Desmet, 2012a; Broekens and Brinkman, 2013; Huisman et al., 2013; Obaid et al., 2015).

Verbal tools enable users to report their current affective state by selecting or rating words that best express their feelings. They are the most commonly used techniques to measure emotional responses to food (Churchill and Behan, 2010; Dorado et al., 2016). However, they have several shortcomings: *(1)* emotions (especially mixed or complex ones) are difficult to verbalize and the labels used to describe them are inherently ambiguous (Köster and Mojet, 2015a), *(2)* emotional terms vary in number and connotation across cultures and languages, (e.g., (Curia et al., 2001; Gutjar et al., 2015b; van Zyl and Meiselman, 2015)), and *(3)* individuals vary in their education level and therefore their vocabulary and general language ability (Prescott, 2017). As a result, emotion words may not be directly equivalent in different languages (Wierzbicka, 1999), and translation of emotion words using dictionaries or backtranslation approaches may miss differences among cultures in emotion intensity, context and other semantics (Boster, 2005b). In addition, verbal tools are demanding for the user since they require cognitive effort (interpretation) and a significant amount of time to fill them out (and this disadvantage increases when they need to be repeatedly applied during an experiment). In addition, it has been argued that it is unclear what information most measures using different emotion words actually provide, and that it is therefore better to focus on measures of core affect such as valence and arousal (Prescott, 2017).

Graphical tools rely on the human ability to intuitively and reliably attribute emotional meaning to (simple) graphical elements (Aronoff et al., 1988; Windhager et al., 2008; Larson et al., 2012; Watson et al., 2012), in particular those linked to facial expressions (Tipples et al., 2002; Lundqvist et al., 2004; Weymar et al., 2011). They allow users to report their feelings in a natural and intuitive way by indicating or rating the (part of a) figure that best represents their current affective state. They do not require the users to verbalize their emotions and can therefore be used by individuals with variable education levels. Graphical self-report tools may also be more effective to measure and express mixed (complex) emotions that are hard to verbalize (Elder, 2018). Since facial expressions are a natural and intuitive way to express emotions (Ekman, 2003) it has been suggested that it may be beneficial to replace the subjective linguistic increments on rating scales by iconic facial expressions (Kaye et al., 2017; Torrico et al., 2018b). Swaney-Stueve et al. (2018) indeed successfully evaluated a 7-point bipolar valence scale labeled with emojis in an online experiment in which children reported their affective responses to different pizza flavors and various real-life experiences. They also mentioned that further research was needed to extend this unidimensional emoji scale into a two-dimensional one that also measures arousal. PrEmo (Product Emotion Measurement Instrument) is a cross-culturally validated self-report instrument that employs an animated cartoon character to measure 12 distinct emotions (Desmet et al., 2000c; Laurans and Desmet, 2012a). Although PrEmo has been applied to measure food-evoked emotions (Dalenberg et al., 2014; Gutjar et al., 2015b; den Uijl et al., 2016b; He et al., 2016a; He et al., 2016b), it was not designed for this purpose and requires users to rate several emotions that have no clear relation to food experiences (e.g., *pride, hope, fascination, shame, fear, sadness*). In addition to being time-consuming, it appears to be more sensitive to valence than to arousal (He et al., 2016a). Another related graphical affective self-report tool is the AffectButton (Broekens and Brinkman, 2013), which allows users to respond their affective stated by moving a mouse controlled cursor to adjust the emotional expression of an iconic face. However, this tool requires the user to successively explore the entire affective space to find the desired expression each time a response is given. In contrast to previous graphical self-report tools, the recently developed and validated EmojiGrid (see next section) was specifically designed for the assessment of food-evoked emotions and provides an instantaneous overview of the affective input space.

The EmojiGrid for the self-assessment of valence and arousal

According to the circumplex model of core affect (Russell, 1980b), emotions are characterized by both their valence (pleasantness: the degree of positive or negative affective response to a stimulus) and arousal (the intensity of the affective response to a stimulus). We recently introduced the EmojiGrid: a emoji-based (graphical) self-report tool to measure food-related valence and arousal ((Toet et al., 2018b); see also: Figure 8-1). Emoji are pictographs representing cartoon-like faces with different emotional expressions. They are widely used in electronic messages and Web pages to supplement or substitute written text (). Emoji can support users to express and transmit their intention clearly and explicitly in computer mediated communication (dos Reis et al., 2018). While people do not easily name food-related emotions, they appear to use emoji in a spontaneous and intuitive way to communicate food-related emotional experiences (Vidal et al., 2016). Emoji span a broad range of emotions, varying in valence (e.g., smiling face vs. angry face) and arousal (e.g., sleepy face vs. excited face). It has been found that emoji with similar facial expressions are typically attributed similar meanings (Moore et al., 2013; Jaeger and Ares, 2017) that are also to a large extent language independent (Kralj Novak et al., 2015). The EmojiGrid enables observers to report their feelings by directly relating them to the facial expressions of emoji. In a previous validation study, we found that the EmojiGrid is self-explaining and intuitive: valence and arousal ratings were independent of the presence or framing of verbal instructions (Toet et al., 2018b). This suggests that the EmojiGrid may be a valuable tool for cross-cultural studies since it is independent of language and requires only minimal instructions ("*Click on the grid*").

Although the relation between subjective valence and arousal ratings depends both on personality and culture at the idiographic level (i.e., within individuals; (Kuppens et al., 2017)), the shape of this relation is typically characterized by a U-shape at the nomothetic or group level (i.e., across persons and for a wide range of different stimuli such as sounds, music, paintings, images, movies, words, facial expressions, odors: (Kuppens et al., 2013; Mattek et al., 2017)). This U-shaped relation between valence and arousal remains the dominant pattern across cultures, although the steepness of the curve varies per country (Kuppens et al., 2017). Cultural preferences for reporting low or high arousal states seem to pull the overall nomothetic U-shape into one direction or another. In general, Western countries tend to show a steeper U-shape than Eastern countries (Kuppens et al., 2017). In a previous validation study with Dutch participants, we found that the mean valence and arousal ratings for food images obtained with the EmojiGrid also reflected this universal U-shaped relation, in the sense that the arousal values monotonously increased from the center of the valence scale towards its extremes (Toet et al., 2018b). Food images that were rated as high (pleasant) or low (unpleasant) on valence were typically rated as more arousing than images rated near neutral on valence. Also, arousal ratings below neutral were scarcely reported, meaning that most food items were typically perceived as stimulating rather than de-activating.

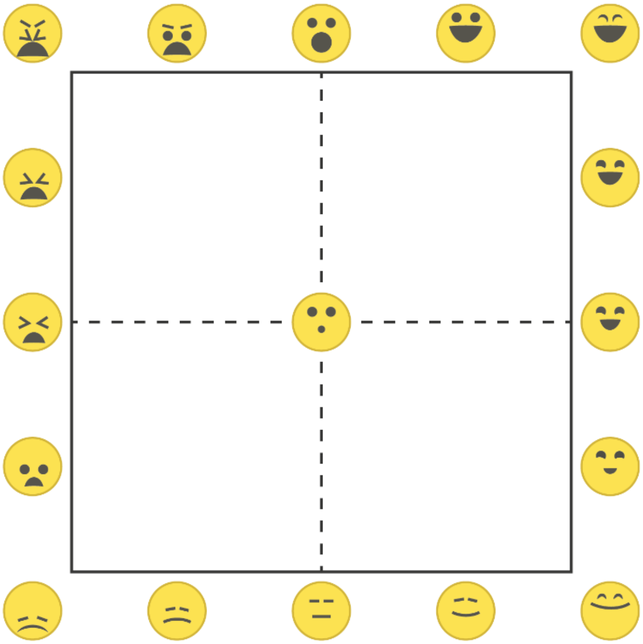


Figure 8-1. The EmojiGrid: an emoji labeled Affect Grid for the measurement of food related affective associations. The facial expressions of the emoji vary from disliking (unpleasant) via neutral to liking (pleasant) along the horizontal (valence) axis, and gradually increase in intensity along the vertical (arousal) axis.

The current study was performed to assess the value of the EmojiGrid for cross-cultural studies. Respondents from three European (GE, NL, UK) countries with different mother languages and one Eastern (JP) country rated valence and arousal for 60 different food images using the EmojiGrid. The images were selected to cover the entire valence range (ranging from positively valenced images showing e.g. fresh fruit and sweets, expected to be of relatively higher valence, to negatively valenced images depicting e.g. molded and spoiled food, expected to be of relatively lower valence). Our first hypothesis is that (H1) a U-shaped relation between valence and arousal is nomothetic for all four groups. In addition, we hypthesize that, compared to the European groups, the JP group shows (H2) a less steep overall U-shape (i.e., the mean arousal ratings of this group depend less strongly on valence; (Kuppens et al., 2017)) and (H3) a less extreme response style, (i.e., the mean valence and arousal ratings of this group cover a smaller range of the response scale; (Yeh et al., 1998; Yao et al., 2003)). To the best of our knowledge there have been no previous cross-cultural studies that investigated the relation between valence and arousal for food stimuli covering the entire valence range (from highly non-appetitive to very appetitive).

Methods

Measures

Demographics and state variables

People's responses to food images are known to be affected both by their current internal state and by their BMI (Body Mass Index; (Burger et al., 2011)). In addition to their age and gender, participants were therefore also asked to report their current physical state (degree of hunger, thirst and fullness), measured with visual analog scales (VAS) labeled from "*very hungry/thirsty/full*" to "*not hungry/thirsty/full at all*". The VAS scales were displayed as solid lines and participants responded by clicking with the mouse on the appropriate location on the lines. Responses were analyzed by converting distances along the lines to scales ranging from 0 to 100, although this was not explicitly displayed to the participants.

Valence and arousal: the EmojiGrid

Valence and arousal were assessed with the EmojiGrid self-report tool (see: Figure 8-1; see also (Toet et al., 2018b)). The EmojiGrid is a Cartesian axes system similar to the Affect Grid (Russell et al., 1989), but the verbal labels on the midpoints and endpoints of the axes are replaced with emoji showing food-related facial expressions. Also, additional emoji are inserted between the midpoints and the endpoints of each axis (resulting in five emoji on each side of the grid), and one (neutral) emoji is placed in the center of the grid, resulting in a total of 17 emoji on the grid. The central emoji with a neutral expression serves as a baseline or anchor point. The facial expressions of the emoji vary from disliking (unpleasant) via neutral to liking (pleasant) along the horizontal (valence) axis, and gradually increase in intensity along the vertical (arousal) axis. The facial expressions are defined by the eyebrows, eyes and mouth configuration of the face, and are inspired by the Facial Action Coding System (Ekman and Friesen, 2003). The arousal dimension is represented by the opening of the mouth and the shape of the eyes, while the valence dimension is represented by the concavity of the mouth, the orientation and curvature of the eyebrows, and the vertical position of these features in the face area (representing a slightly downward looking face for lower arousal values and a slightly upward looking face for higher valence values). These facial features represent a minimal set needed to express the range of emotions over the Affect Grid. To avoid potential biases in ratings due to the emotional connotation of colors (Clarke and Costall, 2008; Suk and Irtel, 2010), we adopted a monochromatic neutral (yellow) color scheme in the design of the EmojiGrid. Users respond by placing a check mark at the location that corresponds to the emoji (facial expression) that best represents their affective state after perceiving a certain food or beverage. A simple instruction asking participants to respond by clicking on the EmojiGrid is sufficient to clarify its use and meaning. The two dimensions of the EmojiGrid (corresponding to the affective dimensions of valence and arousal) were converted to a range from 0 to 100 points. In a previous study (Toet et al., 2018b) we found that the EmojiGrid is a robust, reliable and intuitive affective self-report tool that does not rely on written instructions and that can efficiently be used to measure food-related emotions.

Participants

In this study we tested convenience groups from Germany (GE), the Netherlands (NL), the United Kingdom (UK), and Japan (JP), implicitly adopting country as a (albeit imperfect) proxy for culture.

Dutch participants were recruited through postings on social media and direct emailing. Groups from Germany and the United Kingdom were recruited via the Prolific database (https://prolific.ac), and participants from Japan via the Crowdworks database (https://crowdworks.jp). Exclusion criteria were age (either younger than 18 years or older than 70 years), color vison deficiencies and food allergies. The experimental protocol was reviewed and approved by the TNO Ethics Committee (Ethical Approval Ref: 2017-011) and was in accordance with the Helsinki Declaration of 1975, as revised in 2013 (World Medical Association, 2013). Participation was voluntary. After completing the study, all participants were offered a small financial compensation for their participation.

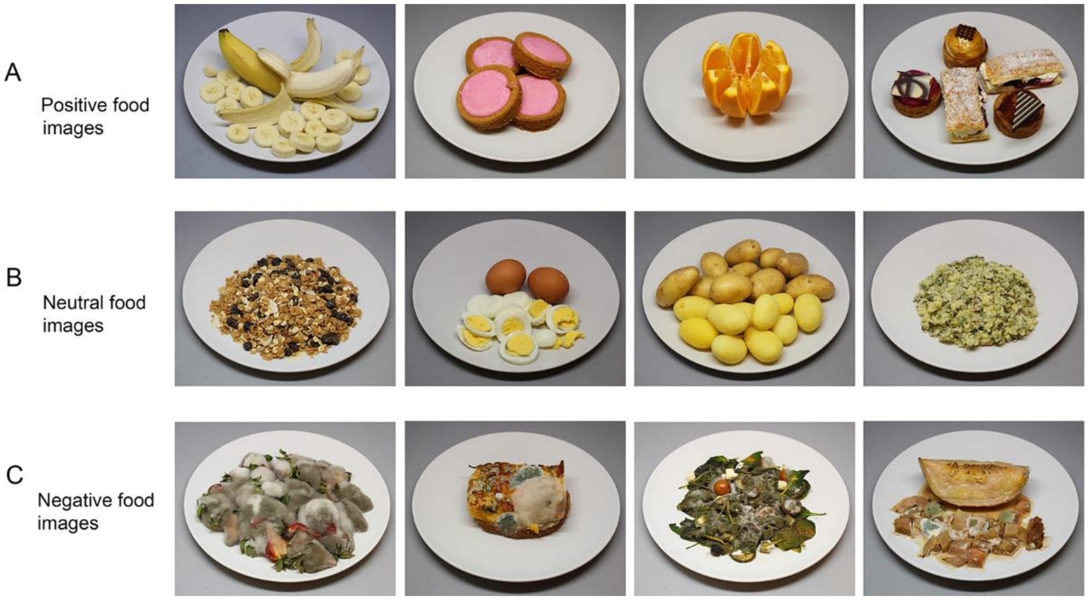


Figure 8-2. Stimulus examples. (A) Positive food images (e.g., banana, cookies, orange, sweets), (B) neutral food images (e.g., cereals, boiled eggs, boiled potatoes, hotchpot) and (C) negative images of rotten food (e.g. strawberries, omelet on bread, Greek salad, melon).

Stimuli

The stimulus set consisted of 60 different food images: 50 images were specifically registered for this study according to a standard protocol (see (Charbonnier et al., 2016); for some examples see: Figure 8-2) and 10 additional images were taken from the FoodCast research image (FRIDa) database ((Foroni et al., 2013); see: Figure 8-3). The 50 images that were registered for this study have a resolution of 1037 × 691 pixels and represent natural food (e.g., strawberry, salad), rotten or molded food (e.g., rotten banana, molded salad), raw food (e.g., raw chicken, raw potatoes), processed food (e.g., cakes, fried fish), unfamiliar food (e.g., locusts), and contaminated food (e.g., hotchpot with fake turd). The set of food items was selected such that their perceived valence is likely to be distributed along the entire scale (ranging from very low valence for rotten, molded or contaminated food, via neutral for raw onions, boiled eggs or potatoes, to very high valence for fresh fruit, chocolates and pastries). The 10 additional food images from the FRIDa database (Figure 8-3) have a resolution of 530 × 530 pixels and were selected such that their associated valence scores (as reported in their accompanying data file, see: (Foroni et al., 2013)) also cover a wide range of the scale. The validated FRIDa images were included as anchor points for verification purposes (see (Toet et al., 2018b)). In the literature most studies on food elicited emotions mainly include appetitive stimuli (e.g., (Blechert et al., 2014c; Miccoli et al., 2014; Miccoli et al., 2016)). In this study we explicitly include non-appetitive stimuli since tools for product-elicited emotion research should also be validated for negative emotions if they are to be applicable to a broad range of foods and beverages (Meiselman, 2015b).

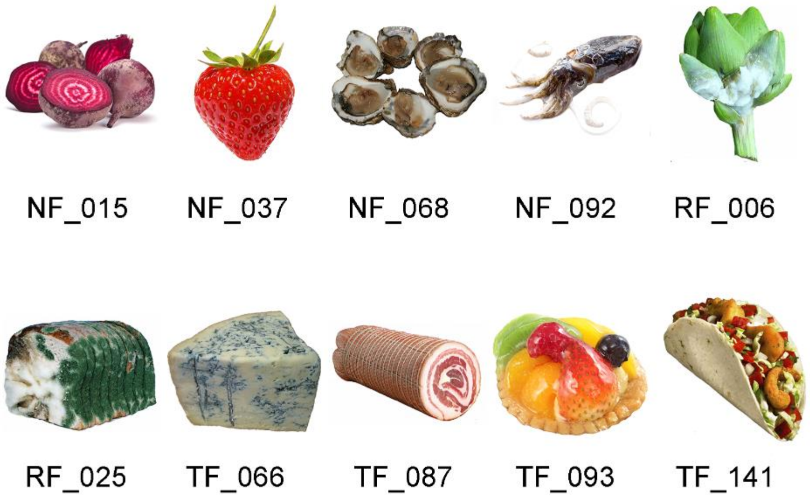


Figure 8-3. The ten stimuli from the FRIDa database (images reproduced with the permission of the copyright holder: (Foroni et al., 2013)). The labels are the original identifiers of the images in the FRIDa database (NF=natural food, RF=rotten food, TF=transformed food). This set consists of two highly negative (RF) and three highly positive (NF_037, TF_093, TF_141) images. The remaining five images are distributed over the neutral zone of the valence-arousal continuum.

Procedure

Participants took part in an anonymous online survey. The survey commenced by presenting general information about the experiment and thanking participants for their interest. Also, the participants were asked to perform the survey on a (laptop) computer (and not on a small handheld device such as a smartphone) and put their web browser in full-screen mode to maximize the questionnaire resolution and avoid external distractions such as software running in the background. Then the participants were informed that they would see 60 different food images during the experiment and they were instructed to rate their first impression of each image without worrying about calories. It was emphasized that there were no correct or incorrect answers and that it was important to respond seriously. Subsequently, participants electronically signed an informed consent by clicking "*I agree to participate in this study*", affirming that they were at least 18 years old and voluntarily participated in the study. The survey then continued with an assessment of the demographics (age, gender, height and weight) and the current physical state (degree of hunger, thirst and fullness) of the participants.

Next, the participants were shown the EmojiGrid response tool together with an explanation about how they should use the tool to report their (valence and arousal) rating for each image. The instructions merely stated: "*Click on a point in the grid that best matches your feelings towards the picture*". No further reference was made to the dimensions of valence and arousal. Then they performed two practice trials to further familiarize them with the use of this tool. Immediately after these practice trials, the actual experiment started. The 60 different food images were randomly presented over the course of the experiment. The entire experiment typically lasted about 10 minutes on average.

Data analysis

IBM SPSS Statistics 25 (www.ibm.com) for Windows was used to perform all statistical analyses. Intraclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated based on a mean-rating ($k = 3$), absolute-agreement, 2-way mixed-effects model (Shrout and Fleiss, 1979; Koo and Li, 2016). ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, while values greater than 0.90 indicate excellent reliability (Koo and Li, 2016). For all other analyses a probability level of $p < .05$ was considered to be statistically significant.

For each of the four countries investigated (GE, JP, NL, UK), we computed the mean valence and arousal responses for each image across all participants. Matlab 2018a (www.mathworks.com) was used to investigate the relation between the (mean) valence and arousal ratings for each country and plot the data. The Curve Fitting Toolbox (version 3.5.7) in Matlab was used to compute a least-squares fit of a quadratic function to the data points.

Results

There was no indication of systematic responding (i.e., systematically using the same part of the grid), so no responses were excluded. Descriptive data for each stimulus across evaluative dimensions (means, standard errors) and groups are available as supplementary material.

Demographics and state variables

Table 8-1 lists the demographic data for the four groups tested in this study. The age of the participants ranged between 18-65 years. One-way analysis of variance showed a significant difference in mean age between the four groups $F(3,281)=15.04$, $p<0.001$. However, the actual difference in mean age was only 10 years, while the 95% confidence intervals for age ranged between 28 and 40 years for all groups. Thus, participants were not in an age range that is likely to affect their ratings (Boyce and Shone, 2006; Kremer et al., 2007).

Table 8-1. Demographics of the four different groups.

Group	N	Male	Female	Age	
				M (SD)	95% CI
GE	63	31	32	30.5 (9.2)	[28.2,32.8]
JP	99	33	66	37.2 (10.2)	[35.2,39.3]
NL	62	38	24	27.2 (11.3)	[24.3,30.0]
UK	61	14	47	36.5 (11.5)	[33.6,39.5]

Table 8-2 lists the measures describing the physical state of the four groups tested in this study There was a significant difference in BMI between the four groups F(3,281)=13.28, p<0.001. Post-hoc comparisons using the Bonferroni correction indicated that the BMI differed significantly between GE and JP, between NL and UK, and between UK and JP. The 95% confidence intervals show that the GE, NL, and JP groups are within the normal BMI range (i.e., between 18.5 and 25), while the UK group was slightly overweight. Participants indicated low hunger and thirst levels, and only medium fullness levels. One-way analyses of variance showed no significant difference between the groups in hungriness, thirstiness and fullness. Based on these self-reports, participants were not in a physiological state that could constrain their ratings. Also, there was no significant correlation between BMI, hungriness, fullness or thirstiness and the dependent variables valence and arousal. Hence, these state variables were not included as covariates in the further analysis.

Data reliability

To quantify the agreement between the mean ratings provided by participants from the different countries, we computed intraclass correlation coefficient (ICC) estimates between all groups, both for the mean valence and arousal ratings. Figure 8-4 shows that the ICC values for the valence ratings range between 0.92 and 0.99, indicating excellent reliability (even though it was an internet experiment and we couldn’t control for many factors as in a lab experiment). Figure 8-5 shows the ICC estimates for the mean arousal ratings. Results are shown for all food images (overall) and separately for images with either low (below neutral) or high (above neutral) mean valence ratings. This figure shows that there is a moderate to good interrater agreement across all groups tested for images with high mean valence ratings (ICC values ranging from 0.68 to 0.85). For images with low mean valence ratings the interrater agreement is excellent between the different European groups (ICC values ranging from 0.96 to 0.97). However, the agreement between the mean ratings from Japanese and European participants is poor for images with low mean valence ratings (ICC values ranging from 0.23 to 0.27). As a result, the overall agreement between Japanese and the European raters is also poor (ICC values ranging from 0.29 to 0.36). This result agrees with a previous finding that for negative images, Asian arousal scores were lower compared to those of Western participants, while there was no significant difference for positive and neutral images (Torrico et al., 2018b).

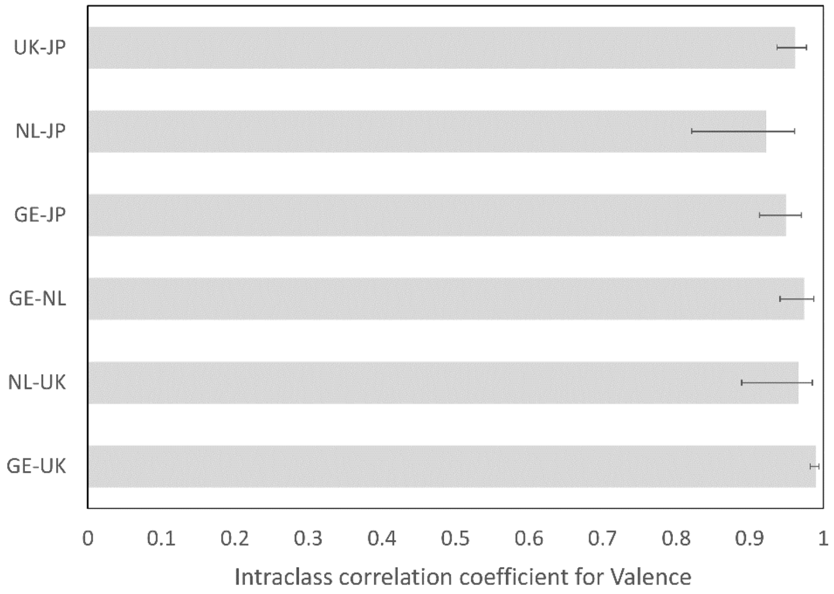


Figure 8-4. Intraclass correlation between the mean valence ratings for each of the four different nations (GE, JP, NL, UK). Error bars represent the 95% confidence intervals.

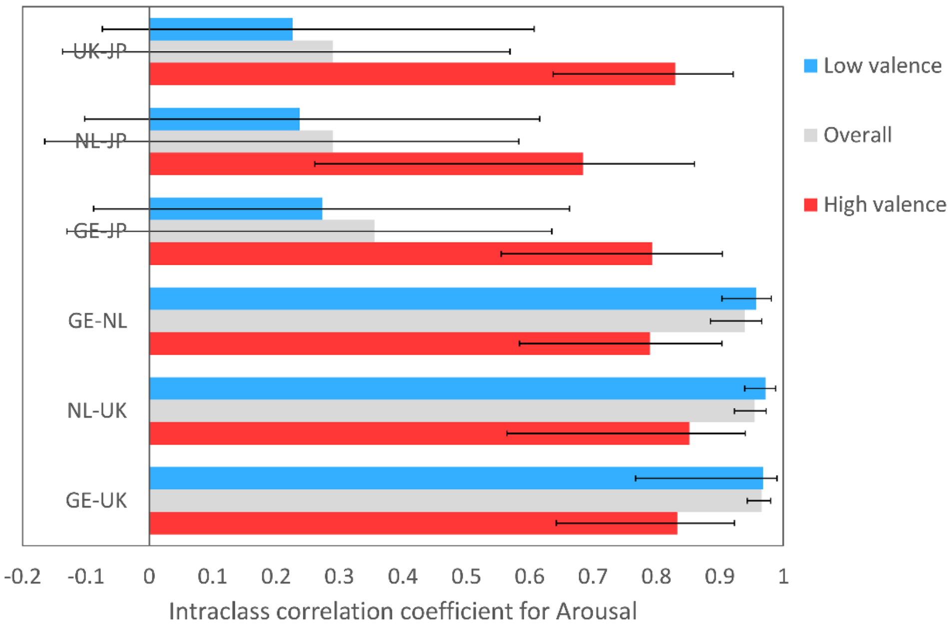


Figure 8-5. Intraclass correlation between the mean arousal ratings for each of the four different nations (GE, JP, NL, UK). Results are shown for all food images (overall) and separately for images rated as being either low or high-valence.

Valence and arousal

Figure 8-6 shows the relation between the mean valence and arousal ratings for all 60 stimuli and for the four different cultures tested (GE, JP, NL, UK). The curves represent least-squares quadratic fits to the data points. The adjusted R-squared values represent the agreement between the data and the quadratic fits, and range between 0.83 and 0.89, indicates excellent fits. This figure shows that the nomothetic (group level) relation between valence and arousal displays a U-shape for each of the four groups tested, confirming our first hypothesis (H1). In addition, it appears that the nomothetic relation between valence and arousal closely agrees between the three European groups (GE, NL, UK). Japanese participants tend to rate low-valence stimuli as less arousing than European participants, while they tend to rate high-valence stimuli as somewhat more arousing. A polynomial regression with arousal as dependent and valence and country as independent variables showed that the average arousal score for Japan differs significantly from that of the UK, Germany and the Netherlands (Table 8-4). Similar analyses with each of the European countries as reference category revealed no significant differences between any of them. No assumption violations were found. Thus, our hypothesis (H2) that (compared to the European groups) the JP group shows a less extreme (MRS) response style is only partly (i.e., only for low-valenced stimuli) confirmed.

For each of the four groups tested, we computed the range of the scales covered by the mean valence and arousal ratings. The results (listed in Table 8-3) show that, compared to the European groups, the mean valence and arousal ratings provided by the JP group covers a smaller fraction of the scales, thus confirming our third hypothesis (H3). Closer inspection of the data (as illustrated in Figure 8-7) reveals that all four groups provided the highest mean valence ratings for fruit and sweets and the lowest mean valence ratings for molded food, in line with previous studies (Foroni et al., 2013; Blechert et al., 2014c; Miccoli et al., 2016; Prada et al., 2017).

Table 8-2. Physical state of the four different groups.

Group	BMI		Hungriness		Thirstiness		Fullness	
	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
GE	23.9 (4.3)	[22.8,25.0]	29.2 (28.6)	[22.0,36.4]	42.0 (24.9)	[35.8,48.3]	48.2 (25.6)	[41.8,54.7]
JP	21.4 (3.4)	[20.7,22.0]	28.4 (23.7)	[23.7,33.2]	39.6 (24.4)	[34.7,44.5]	48.1 (21.8)	[43.7,52.4]
NL	22.4 (3.1)	[21.6,23.2]	32.5 (21.9)	[26.9,38.1]	46.8 (21.9)	[41.3,52.4]	54.9 (23.2)	[49.0.60.8]
UK	25.6 (6.3)	[24.0,27.2]	34.5 (24.7)	[28.2,40.9]	44.4 (25.5)	[37.9,50.8]	52.1 (26.1)	[45.4,58.8]

Table 8-3. Fraction (%) of the valence and arousal scales covered by the mean ratings for each of the four groups tested.

	GE	JP	NL	UK
Valence	78.6	68.1	80.3	76.2
Arousal	43.7	32.1	50.4	43.7

Figure 8-8 shows the relation between the mean valence and arousal ratings for each of the 60 individual food images tested and between all pairs of the four different cultures investigated (GE, JP, NL, UK). This image shows that the mean ratings for both high- and low-valence images are quite similar (i.e., shorter arrows) between the different European groups (GE-NL, GE-UK, NL-UK, top row in Figure 8-8), while images with intermediate mean valence ratings show somewhat larger and irregular variation in ratings (i.e., longer arrows). However, the relation between the mean ratings provided by JP group and by each of the European groups (bottom row in Figure 8-8) shows a quite systematic variation. Japanese participants systematically give lower mean arousal ratings to images with below-neutral (i.e. < 50) mean valence ratings and also systematically give lower mean valence ratings to images with above neutral (i.e. > 50) mean valence ratings.

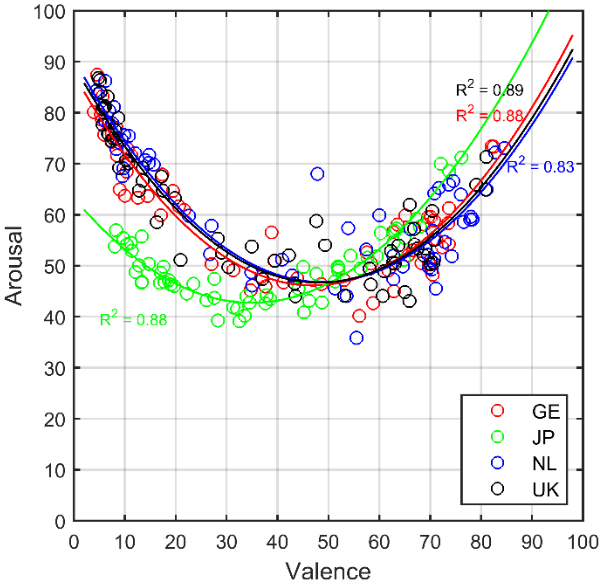


Figure 8-6. Relation between the mean valence and arousal ratings for the four different cultures tested (GE, JP, NL, UK). The curves represent least-squares quadratic fits to the data points. The adjusted R-squared values represent the agreement between the data and the quadratic fits.

To investigate the differences in food evoked affective feelings across the different countries, we computed for each food image and each of the six country pairs the Euclidian (vector norm) distance between the mean valence and arousal ratings (i.e., the length of the vectors shown in Figure 8-8). Then, we computed the average distance over all food items for each country pair (see: Figure 8-9). It appears that the average (overall) rating distance is lowest between the GE and UK groups (i.e., they have the highest similarity), while the NL group has a similar distance to the GE and UK groups. The distance between the JP and the European groups is more than twice the distance between European group pairs.

Table 8-4. Polynomial regression with arousal as a function of valence and country.

Variables	B	SE
Valence	-1.604***	0.076
Valence^2	0.017***	0.001
Country-pair_GE-JP	6.277***	1.148
Country-pair_NL-JP	7.219***	1.184
Country-pair_UK-JP	6.982***	1.144

Notes: Sample size = 240. R2=0.755, F = 144.568***. ***p<0.001

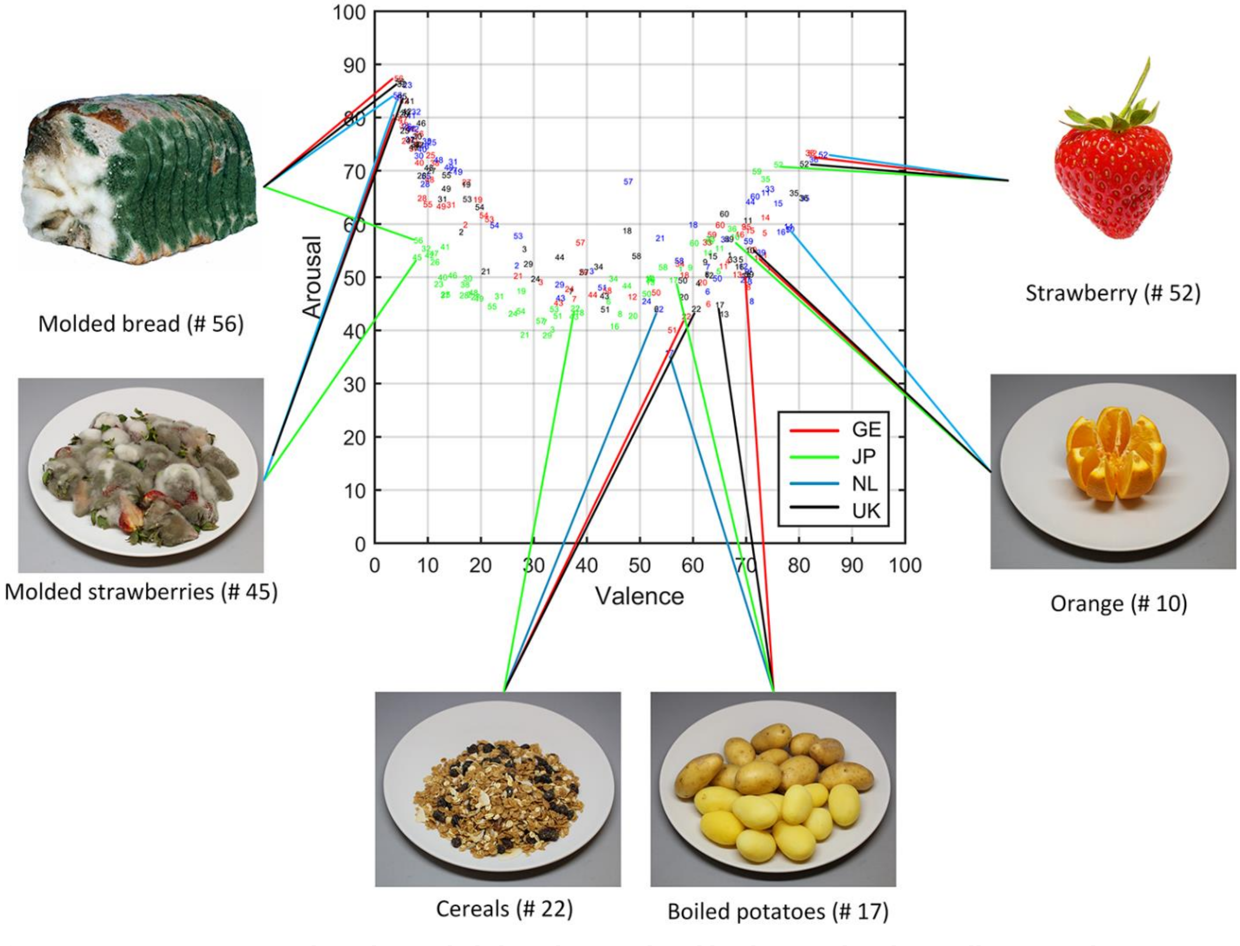


Figure 8-7. As Figure. 8-6 (where the symbols have been replaced by the stimuli indices), illustrating the mean ratings for two appetitive (# 10 and 52), two neutral (# 17 and 22) and two non-appetitive (# 45 and 56) stimuli for each of the four groups tested (GE, JP, NL, UK). Images # 52 and 56 are reproduced with the permission of the copyright holder (Foroni et al., 2013).

Discussion

In this study we investigated whether the graphical and language-independent EmojiGrid can serve as an affective self-report tool for cross-cultural studies. In an online experiment groups from Germany (GE), Japan (JP), the Netherlands (NL) and the United Kingdom (UK) rated valence and arousal for 60 different food images (covering a large part of the affective space) using the EmojiGrid. The results confirm our first hypothesis (H1) that the nomothetic relation between valence and arousal has the well-known U-shape for all groups. The European groups (GE, NL and UK) closely agree in their overall rating behavior. Compared to the European groups, the Japanese group systematically gave lower mean arousal ratings to images with mean valence ratings below neutral and systematically gave lower mean valence ratings to images with mean valence ratings above neutral. These results partly (i.e., only for low-valenced stimuli) confirm our second hypothesis (H2) that, compared to the European groups, the JP group shows a less steep overall U-shape. Our results also confirmed our third hypothesis (H3) that, compared to the European groups, the JP group used a smaller range of the affective rating scales. Since we included a relatively large number of food items at the low valence end of the scale (which is not common in food research), and since we measured arousal instead of hedonic ratings, we were able to uncover a more specific relation for this general finding. The specific way in which JP respondents utilized a smaller range of the scale depended on whether the presented food items were generally rated as pleasant, in which case they used a smaller range of the valence scale, or unpleasant, in which case they used a smaller range of the arousal scale. Overall, our present findings reflect the MRS response style that is typically observed for JP groups (Chen et al., 1995; Lee et al., 2002; Lottridge et al., 2012), and agrees with previous findings that JP participants typically use a smaller range of hedonic scales (Yeh et al., 1998; Yao et al., 2003).

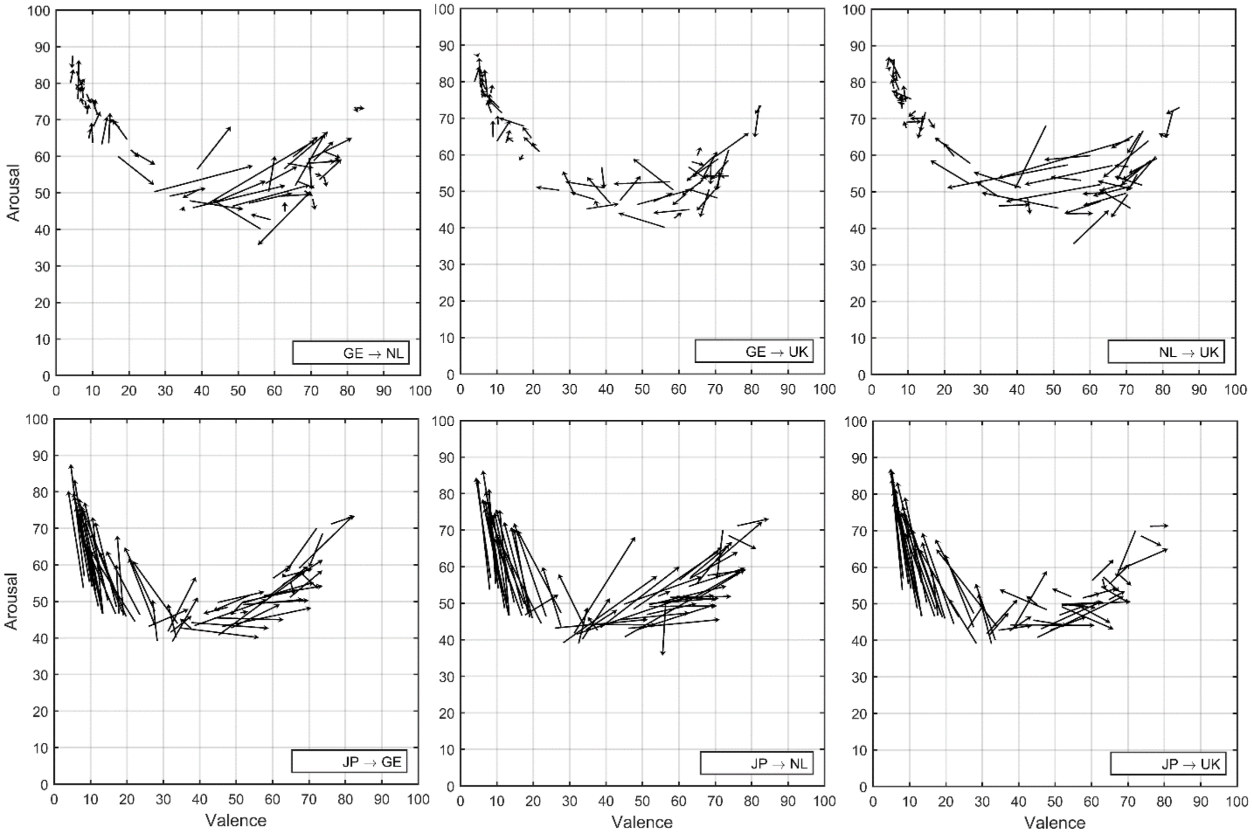


Figure 8-8. Relation between the mean valence and arousal ratings for each of the 60 individual food images tested and between all six pairs of the four different groups investigated (GE-NL, GE-UK, NL-UK, JP-GE, JP-NL, JP-UK).

Our current results agree with previous observations that emoji can serve as a direct self-report tool for measuring food related affective feelings (Vidal et al., 2016; Ares and Jaeger, 2017; Gallo et al., 2017; Jaeger et al., 2017c; Jaeger et al., 2018a; Schouteten et al., 2018). It appears that users easily connect emoji to food-elicited emotions, even without any explicit reference to feelings in the wording of the associated question (Ares and Jaeger, 2017). The expressions of the emoji along the axis of the EmojiGrid systematically vary in valence and arousal and can easily be related to food-evoked emotions. In addition, the two-dimensional layout of the grid provides instantaneous access to the entire valence and arousal space, making the EmojiGrid an efficient response tool (a single click suffices to respond both valence and arousal). In contrast, most previous studies used emoji that were simply selected as the most appropriate ones from the general set of currently available emoji and that were not specifically designed to measure food-related emotions. As a result, several emoji were obviously out of context and had no relevance for the description food-related affective associations (Jaeger et al., 2017c). Also, the size of the emoji sets used in these studies (33 emoji: (Ares and Jaeger, 2017; Jaeger et al., 2017c; Jaeger et al., 2018a; Schouteten et al., 2018), 25-39 emoji: (Jaeger et al., 2017a), and 50 emoji: (Gallo et al., 2017)) was overwhelming and comparable to the large number of words typically used in emotional lexicons to measure emotional associations to food and beverages (e.g., (King and Meiselman, 2010; Spinelli et al., 2014; Nestrud et al., 2016a)). For repeated or routine testing in applied settings, selecting emoji from a long list of possible candidates is obviously too demanding, and a tool like the EmojiGrid will be more practical.

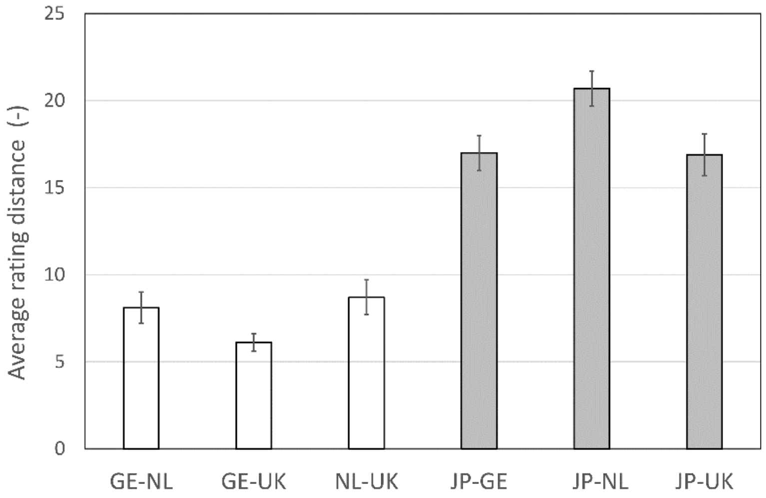


Figure 8-9. Average (over all food items tested) distance between valence-arousal ratings for each group pair. Error bars represent the standard error of the mean.

In this study we did not include participants younger than 18 years. However, it is likely that our findings will also apply to young people since it has been found that both the use of emoji (Nishimura, 2015) and their interpretation (Jaeger et al., 2018c) are independent of age. Their intuitive visual display of emotion also makes emoji particularly suitable both for use with children who may not have the vocabulary to convey all their emotions (Gallo et al., 2017; Schouteten et al., 2018; Swaney-Stueve et al., 2018) and with individuals with variable education levels (Prescott, 2017). Given that emotions in facial expressions, gestures and body postures are similarly perceived across different cultures (Ekman and Friesen, 1971; Ekman, 1994), cross-cultural differences in the interpretation of emoji may also be smaller than the influences of culture and language on verbal affective self-report tasks (Jaeger and Ares, 2017; Torrico et al., 2018b).

To further our understanding of specific product elicited emotions, we suggest that future studies combine the EmojiGrid with biometric measurements (e.g., facial expressions or physiological responses like heart rate and galvanic skin conductance) as additional and more objective indicators of valence and arousal.

We conclude that the EmojiGrid appears to be a valid and language-independent affective self-report tool for cross-cultural research on food-related motions, that reliably reproduces the familiar nomothetic U-shaped relation between valence and arousal across cultures, with shape variations reflecting established cultural characteristics. Next to the affective assessment of food, the EmojiGrid may also be a useful tool to study the emotional response to other sensory (e.g., olfactory, tactile, auditory, visual) stimuli. The implementation on mobile devices may be used to afford the real-time affective evaluation of environments or large-scale events.

Acknowledgements

The authors thank Jef van Schendel for designing the emoji for the EmojiGrid, and Robin van Stokkum for his help with the polynomial regression.

Supplemental materials

The EmojiGrid, the 60 food images used as stimuli in the experiments, and an Excel file with the results are all available from the Figshare Repository:
https://figshare.com/s/1956bda3843f93a20077

Chapter 9

Comparing Explicit and Implicit Measures for Assessing Cross-Cultural Food Experience

This chapter is published as:
Kaneko, D., Stuldreher, I., Reuten, A.J., Toet, A., Van Erp, J.B.F., and Brouwer, A.M. (2021). Comparing explicit and implicit measures for assessing cross-cultural food experience. Frontiers in Neuroergonomics 2, 5.

Abstract

The present study investigated the potential of implicit physiological measures to provide objective measures of affective food experience in contrast to explicit self-report ratings in a cross-cultural context. Dutch and Thai participants viewed 120 food images portraying universal food image categories (regular and molded food) and cultural food image categories (typically Dutch and Thai food). The universal food images were taken as ground truth high and low valence stimuli, where we assumed no genuine difference in affective experience between nationalities. In contrast, for the cultural food images, we did expect a genuine difference between nationalities. Participants were asked to rate valence, arousal and liking of each food image. In addition, heart rate (HR) and phasic electrodermal activity (EDA) responses to the images were recorded. Typically Asian and Western response biases were found for explicit ratings of regular and molded food with an extreme response style for Dutch, and a middle response style for Thai participants. However, such bias was not observed in HR. For cultural food image categories, HR showed the hypothesized interaction between participant nationality and food image category, reflecting the expected genuine difference between nationalities in affective food experience. Besides presenting participants with images, we also asked participants to taste typically Thai and Dutch drinks. Similar to images, a significant interaction between participant nationality and cultural food category was found for HR. An interaction was also found for sip size, while this was not seen in explicit measures. We attribute this to differences in the moment that these measures were taken. In this study, phasic EDA did not appear to be a sensitive measure of affective food experience, possibly since stimuli mostly differed in valence rather than arousal. To conclude, our study constitutes an example where cultural bias negatively affected the accuracy of self-reports, and only the implicit physiological measures followed the prior expectations of genuine food experience, indicating the potential of these measures to study cross-cultural food experience.

Introduction

To predict whether consumers will choose a certain food product, their emotional response when experiencing this product is considered to be an important predictor (Dalenberg et al., 2014; Gutjar et al., 2015b; Köster and Mojet, 2015b; Samant et al., 2017). With the current globalizing trend, measuring these emotions cross-culturally is important for international food marketers (Rozin, 1988; 2006; Meiselman, 2013b; Meiselman, 2015a).

Assessment of food-evoked emotions is predominantly based on explicit measures (‘conscious’ self-report ratings) rather than on implicit measures (‘unconscious’ physiological and behavioral measures) (Lagast et al., 2017; Kaneko et al., 2018a). Explicit measures are relatively easy to apply, practical for quantitative analysis, and widely believed reliable for most sensory and psychological studies (Lawless and Heymann, 2010; Dorado et al., 2016). Many large international companies explore food product-elicited emotions across different countries by simply translating emotion questionnaires into multiple languages (Meiselman, 2015a). However, the drawback of using explicit self-report questionnaires for assessing affective experience cross-culturally is that cultural background influences how people ‘self-report’, what emotional language they use, and how they use rating scales to describe their own food-evoked emotions (van Zyl and Meiselman, 2015; Silva et al., 2016; van Zyl and Meiselman, 2016a; Ares, 2018). A review by Meiselman (2015a) questioned whether simple translations between languages capture the local meaning of emotional words in a questionnaire well enough - people raised in different cultures may not experience the same emotions evoked by the same stimuli, and evoked emotions may not be expressed in the same manner. For example, Uchida and Kitayama (2009) analyzed American and Japanese descriptions of ‘happiness’ and showed that Japanese associate happiness with ‘social harmony’ while Americans associate it with positive experience of personal achievement. Concerning culture-dependent use of rating scales, Western respondents have been found to have an ‘Extreme Response Style’ (ERS) (using the extremes of rating scales), whereas Asian respondents more often use a ‘Middle Response Style’ (MRS) (using the neutral part of the scale) (Chen et al., 1995; Harzing, 2006; Kaneko et al., 2019b). Problems with translating between languages, and intercultural differences in terms of using rating scales, could potentially be overcome by implicit measures, which reflect fast, non-conscious, and uncontrollable responses (Soto et al., 2005; Lagast et al., 2017; Ares, 2018; Kaneko et al., 2018a).

Several physiological measures have been studied in the context of probing affective experience when tasting and viewing food or food images. Among them, heart rate (HR) and electrodermal activity (EDA) are the most often used implicit physiological measures in recent consumer research. These measures have been shown to distinguish between tasting different beverages, chocolates, liked and disliked food (de Wijk et al., 2012; Danner et al., 2014a; Torrico et al., 2018a; Kaneko et al., 2019a). Outside the food domain, several studies have compared implicit physiological responses to different types of stimuli between individuals from different cultures. One study showed weaker electrodermal responses to disgust-eliciting film clips in Asian-American compared to European-American participants (Soto et al., 2016). No difference between cultural groups was found for physiological responses to stimuli such as acoustic startle (Chinese-American and Mexican-American groups; (Soto et al., 2005)), emotional films (Chinese-American and European-American groups; (Tsai et al., 2000)), and reliving of intense emotional episodes (Hmong-American and European-American groups; (Tsai et al., 2002)). These results may be taken to mean that Asian-Americans and other Americans ‘really’ differed in emotional experience when watching disgust-eliciting movies, and not when experiencing the other types of stimuli.

To the best of our knowledge, only two studies in the food domain used implicit measures to investigate psychophysiological effects of food products between different nationalities (Asian who spent less than two years in Australia and Australian groups; (Torrico et al., 2018a; Torrico et al., 2019)). These studies used rating scales as well as a camera to monitor heart rate, skin temperature and facial expressions to investigate cross-cultural effects of viewing universal and culture-specific food (Torrico et al., 2018a; Torrico et al., 2019). Their results indicated that rated food liking is positively correlated to familiarity, and that skin temperature differentiates between cultural groups when tasting culture-specific food samples (Torrico et al., 2019) while no physiological differences between cultural groups exist when tasting (universal) chocolate samples (Torrico et al., 2018a). These results are in line with the idea that physiological measures reflect the ‘true’ emotion: cultural groups do not differ when tasting universal stimuli, and they do differ (in skin temperature) when tasting samples that are expected to genuinely elicit different emotions. No physiological effects were found besides skin temperature, but we should note that camera-based analysis is usually less precise and suffers more from artifacts caused by (chewing) movements, head orientation, and lighting conditions compared to traditional sensors (Kranjec et al., 2014; Bach et al., 2015; Hassan et al., 2017).

In the present study we investigated whether implicit physiological measures (HR and phasic EDA recorded using traditional sensors) can contribute to comparing affective food experiences across cultures objectively without cultural response biases that affect explicit self-report methods. We compared explicit and implicit responses between two cultural participant groups, Dutch (representative for ERS) and Thai (representative for MRS), toward universal food images (regular and molded food) and cultural food images (typically Dutch and Thai food). We selected universal and cultural food image categories so that we could assume a genuine difference in emotional experience between Dutch and Thai participants for the two types of cultural foods, but no genuine difference for the two types of universal foods. For the latter category, we can safely assume a ground truth affective experience of low valence (unpleasant) and high arousal for the molded compared to the regular food images, in both Dutch and Thai participants. A lack of effect of universal food category would therefore indicate that the measure is insensitive. No differences between the two nationalities are expected for implicit measures for universal food image categories while we expect differences on explicit measures due to culturally-dependent response bias. On the other hand, we expect response differences between Dutch and Thai participants on both explicit and implicit measures for the cultural food image categories. Viewing food images happens in real life, and is practical in (psychophysiological) experiments in which a large number of trials is desirable. However, viewing food images is expected to elicit different responses than being confronted and tasting real food (de Wijk et al., 2012). To extend our study beyond viewing food images, we also examined implicit and explicit affective responses in Dutch and Thai participants when tasting typically Thai and typically Dutch drinks.

Our specific hypotheses are as follows:

1) Nationality affects explicit measures for universal food images (regular and molded food) due to a culturally dependent response bias. ERS is expected for Dutch participants, and MRS for Thai participants.
2) Nationality affects explicit measures for the cultural food images due to a genuine difference in affective experience caused by a difference in familiarity with the types of food (in addition to a culturally dependent response bias), resulting in an interaction between participant nationality and cultural food category.
3) Implicit physiological measures are affected by universal food image category (regular versus molded food), but the effect is the same for both nationalities (i.e. no interaction between participant nationality and universal food image category).
4) Implicit physiological measures toward cultural food image categories (Dutch and Thai food) reflect genuine differences in affective experience between participant groups, resulting in an interaction between participant nationality and cultural food image category.
5) Similar to food images, explicit and implicit responses to tasting cultural drinks show an interaction effect between participant nationality and cultural drink category.

Methods

Participants

42 Thai participants were recruited from Chulalongkorn University in Thailand and 45 Dutch participants were recruited from the participant pool of the research institute where the main part of the research was conducted (TNO Soesterberg, The Netherlands). The recruitment process excluded people with color vision deficiencies; food allergies; diets such as vegetarian, vegan or religion-related; an immigration background; an eating disorder diagnosed in the last three years. Also, people who had visited The Netherlands (for Thai participants) or Thailand (for Dutch participants) and who had lived abroad for more than one month could not participate. Participants were asked not to eat for 1 hour before testing. The experimental protocol was approved by the TNO Institutional Review Board (Ethical Approval Ref: 2019-033) and was in accordance with the revised Helsinki Declaration (World Medical Association, 2013). All participants signed an informed consent sheet before the experiment started and received a reward to thank them for participating in the study after completing the experiment.

Materials

Food images

Food images were selected from the Cross Cultural Food Image Database (CROCUFID) (Toet et al., 2019c), which is a collection of food and non-food images, photographed on a standardized plate using a standardized photographing protocol (Charbonnier et al., 2016). From this database, we selected 60 'universal' food images which are expected to be familiar to participant of both nationalities (47 regular food images and13 molded images); and 60 'cultural' food images (30 typically Dutch and 30 typically Thai food images). As shown in Figure 9-1, the national flag of the food's origin was presented on the right bottom of each image to ensure participants recognizing and interpreting the food in a similar way. Universal dishes were accompanied by an image of a globe.

Figure 9-1. Stimulus examples of (A) universal molded food images (melon and strawberries) and universal regular food images (cucumber and French fries), (B) Dutch food images (herring, wheat bread, liquorice candy, stroopwafels), and (C) Thai food images (dragon fruit, seaweed chips, som tam, grass jelly).

Drinks

As a typically Dutch drink, a popular long seller yogurt drink, Fristi (Friesland Campina B.V. the Netherlands), was used. As a typically Thai drink we selected a Chrysanthemum tea drink (Vitasoy International Holdings Limited, Thailand). Both drinks were served in white plain cups.

Rating scales and implicit behavioral measure

The following rating scales were used to rate emotions evoked by viewing food images and tasting drinks, and to check for the familiarity of the participants with the stimuli:

EmojiGrid: An intuitive visual self-report tool that has been specifically developed for the assessment of food-evoked emotions and that has shown to be suitable for cross-cultural testing (Toet et al., 2018a; Kaneko et al., 2019b). Participants report their emotion by clicking the appropriate location in the grid, where each location is associated with a valence and arousal score, ranging from 0 (lowest) to 100 (highest). The Emojigrid is depicted in Figure 9-2.

Hedonic liking scale: 9-point scale with anchors for each point. The anchors are: 1) “dislike extremely”, 2) “dislike very much”, 3) “dislike moderately”, 4) “dislike slightly”, 5) “neither like nor dislike”, 6) “like slightly”, 7) “like moderately”, 8) “like very much”, and 9) “like extremely” (Lim, 2011).

Familiarity scale: 5-point scale with anchors for each point. Anchors of this five-point scale were labeled: 1) “I do not recognize it”, 2) “I recognize it, but I have not tasted it”, 3) “I have tasted it”, 4) “I occasionally eat it”, 5) “I regularly eat it” (adapted from Tuorila et al. (2001)). This scale was used to check whether Thai and Dutch cultural food images were more familiar to participants of the matching nationality compared to the other nationality.

A scale was used during the tasting session to measure sip size of both cultural drinks by weighing the drink before and after the participant had taken a sip, following the procedure in a previous study (Kaneko et al., 2019a).

Physiological Recording Equipment (Electrocardiogram and Electrodermal Activity)

Electrocardiogram (ECG; for heart rate (HR)) and electrodermal activity (EDA; for phasic EDA) were recorded using an Active Two MkII system (Biosemi B.V., Amsterdam, the Netherlands), with a sampling frequency of 512 Hz. ECG electrodes were placed on the right clavicle and on the lowest floating left rib. EDA was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. Two reference electrodes were attached to the temporal bone behind the ears.

Experimental design and procedure

After participants arrived at the laboratory (depending on the nationality, either located at Chulalongkorn University, Bangkok, Thailand; or TNO Soesterberg, the Netherlands), they were told that the experiment consisted of a ‘tasting session’, a ‘viewing session’, and a ‘rating familiarity session’. The experimenter also explained that ECG and EDA sensors would be attached to measure HR and EDA during the experiment. Participants signed the informed consent form and were seated in a comfortable chair in front of an experimental presentation notebook. Then, the ECG, EDA, and reference electrodes were attached, and all signals were checked. HR and EDA were recorded during the tasting and viewing sessions. A schematic image of the study is shown in Figure 9-2. The total duration of the experiment was approximately 75 minutes.

Tasting session

The procedure of the tasting session followed that used in a previous study (Kaneko et al., 2019a). Before the tasting session started, the experimenter showed and explained how to take a sip and to put the cup down after the sip, and participants performed a practice trial with water. After this there was time for additional practice or instructions when needed. The testing procedure started with the presentation of the name of the drink on the screen. This was the sign for the experimenter to place the appropriate drink in front of the participant. After 5 seconds, the name of the drink disappeared, which was the sign for the participant to take one sip. After taking the sip, the participant put the cup down, sat still and looked at a blank white screen. Thirty-five seconds after the name of the drink had disappeared from the screen, the EmojiGrid and hedonic scale appeared in successive, randomized order. Order was randomized so that the ratings from the two scales could be compared and would not be confounded by a possible order effect. After rating, the name of the next drink appeared on the screen. This procedure was repeated until three drinks had been served: first water for practice, followed by the Thai and Dutch drinks in counterbalanced order.

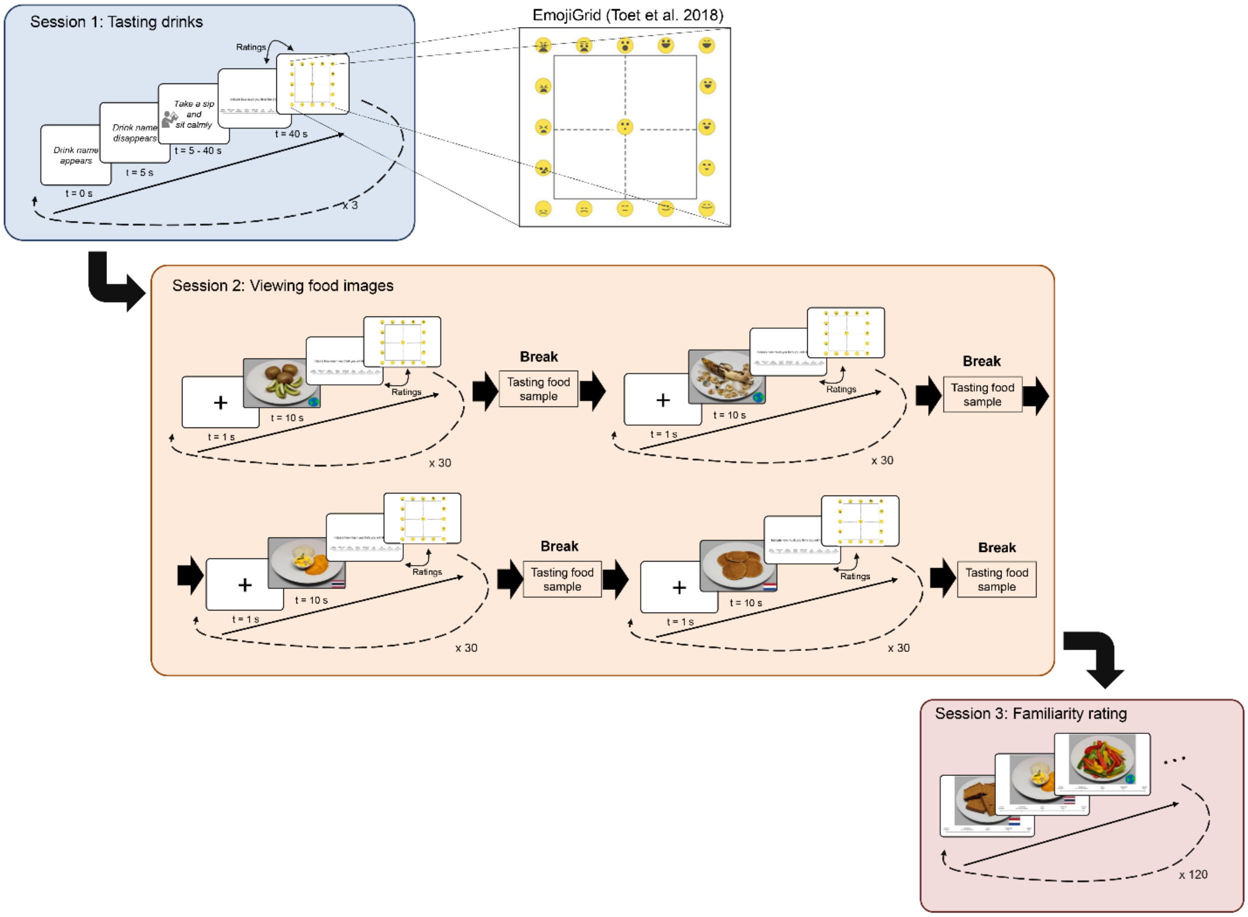


Figure 9-2. Schematic overview of an experimental trial and of the experimental procedure. All participants followed the exact same procedure. Physiological response data were collected during session 1 and 2. At the center, the EmojiGrid (Toet et al., 2018b), used for rating valence and arousal is depicted.

Viewing session

After a short break, the instruction for the viewing session appeared on a screen, and participants had a chance to ask any questions to the experimenter. In this session participants viewed a total of 120 food images in two counterbalanced blocks (universal and cultural food image categories) each consisting of 60 randomized images. A fixation cross was presented for 1 s, after which a food image was presented for 10 s. After 10 s of viewing time, the EmojiGrid and hedonic scale appeared in successive, randomized order. Participants had unlimited time to provide their ratings using the mouse but were instructed to follow their initial impression. In an effort to increase participants' engagement with the images, they were told before starting the viewing session that they would be asked to taste some of the depicted food images they rated during each of four breaks. This part of the instruction was formulated in such a way that the participants were led to believe that they could be asked to eat an unpleasant (molded) food item: “There are four short breaks in this session. During these breaks, we will serve you one of the foods depicted in the images that you just saw. We ask you to taste this food. You are permitted to refuse, but we hope you will taste it.”. Tasting breaks were introduced after each half block (30 images), and tasting samples were sliced banana pieces, peanut chocolate candies, a seaweed chip (typically Thai food item) and a small ‘stroopwafel’ (typically Dutch food item).

Familiarity rating session

After finishing the viewing session, participants were instructed to rate their familiarity with all drinks and all food images they rated in the previous sessions. For drinks, the name of the drinks appeared successively on the screen in random order, accompanied by the familiarity scale. Actual tasting was omitted. Food images also appeared successively in randomized order, accompanied by the familiarity scale. Lastly, participants filled out a short demographic questionnaire, asking about age, gender, height and weight.

Physiological data processing

Data from three Thai participants were discarded due to failure of physiological recordings. Data processing was done using Matlab 2020a software (Mathworks, Natick, MA, USA).

Inter-beat intervals were extracted from ECG following Pan and Tompkins (1985) using a Matlab implementation from Sedghamiz (2014). By inversing the inter-beat interval, a heart rate semi-time series was obtained. Intervals exceeding the absolute threshold of 160 bpm or intervals deviating more than 20% from the previous interval were removed. The heart rate semi-time series was transformed to a regular timeseries at a 10 Hz sampling frequency using a piecewise cubic spline interpolation.

Raw EDA was down-sampled to 10 Hz. Continuous Decomposition Analysis as implemented in the Ledalab toolbox for Matlab was used to separate the tonic (slow) and phasic (fast) components of the EDA (Benedek and Kaernbach, 2010). In further analysis, only the phasic component is considered.

For food images, the pre-processed continuous physiological data were divided in epochs that were time-locked to stimulus onset for each image and each participant. The epoch comprises data ranging from fixation-cross onset to 10 seconds after fixation-cross offset. Response traces were baseline corrected based on the average value of the one second that the fixation-cross was presented.

For drinks, the pre-processed continuous physiological data were divided in epochs ranging from 5 seconds after drink name onset to 35 seconds after drink name offset (which was the sign for the participant to pick up the cup and take a sip), for each drink and each participant. Response traces were baseline corrected based on the average value of the five seconds during drink name onset.

Outliers were detected using Matlab’s ‘isoutlier’ function. Image epochs for which the average physiological response value was more than five median absolute deviations away from the median value across images were removed. This resulted in a removal of 0.06% of the HR data and 8.6% of the phasic EDA data. No outlier detection was conducted for physiological response value for drinks.

Grand-average response traces were obtained for each participant and each of the food image categories (universal food image category (regular and molded), cultural food image category (Dutch and Thai), and cultural drink category (Dutch and Thai)) by averaging over all data traces corresponding to the food image categories of each participant. We then computed the mean HR and mean phasic EDA for each participant and each food image category over the 10 seconds following image onset and over 35 seconds following drink name offset.

Statistical analysis

Statistical analysis was conducted with SPSS ver. 25 (IBM, USA). Two-way mixed ANOVAs were performed on each explicit and implicit measure in response to the universal food images, the cultural food images and the cultural drinks. In all of these analyses, the between-subject factor was participant nationality (Dutch or Thai), and the within-subject factor was respectively universal food image category (regular and molded); cultural food image category (Dutch and Thai); and cultural drink category (Dutch and Thai).

Results

Demographics

Table 1 shows the demographic descriptive for Dutch and Thai participants. The two groups did not differ in age (t (85) = 1.662, p = .100) and BMI (t (85) = 1.022, p = .310) according to Welch’s t-test.

Table 1. Demographic variables for the Dutch and Thai participants.

Cultural group	N	Female	Male	Age	BMI
Dutch participants	45	28	17	21.2 (±1.7)	22.29 (±2.77)
Thai participants	42	24	18	20.6 (±1.5)	21.58 (±3.60)

Familiarity scores

Familiarity ratings are presented in Figure 9-3A and B.

Figure 9-3A shows the expected pattern of Dutch participants rating Dutch food as more familiar than Thai food and Thai participants rating Thai food as more familiar than Dutch food. This pattern is statistically corroborated by a significant interaction between participant nationality and cultural food image category ($F(1, 85) = 1299.52$, $p < .001$, $\eta_p^2 = .939$), and indicates that the selection of cultural food images was appropriate. The two-way mixed ANOVA showed no significant main effect of participant nationality on familiarity scores ($F(1, 85) = .69$, $p = .407$), and a significant main effect of cultural food image category (*F*(1, 85) = 41.19, $p < .001$, $\eta_p^2 = .326$) on familiarity scores, with Dutch food images receiving higher familiarity ratings overall. As a comparison, regular food images from the set of universal food images, received very similar familiarity ratings as food images from the own culture: they received an average score of 4.1 from Dutch participant and an average score of 3.4 from Thai participants.

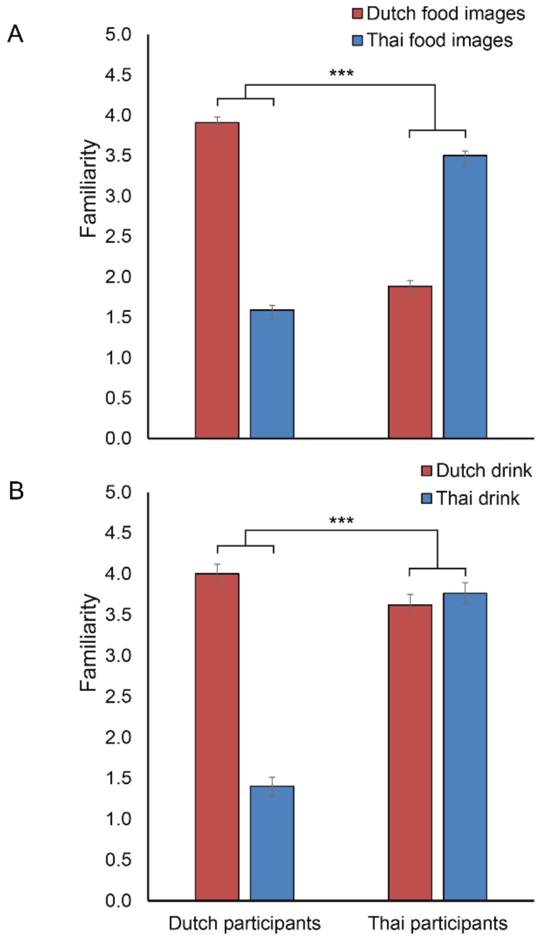


Figure 9-3. Mean familiarity scores of (A) Dutch and Thai food images and of (B) Dutch and Thai drinks rated by Dutch and Thai participants. Error bars indicate standard error of the mean. *** indicates a significant interaction between participant nationality and cultural food category with p < 0.001.

Figure 9-3B and the statistical analysis showed the expected interaction between participant nationality and cultural drink category (*F*(1, 85) = 147.88, $p < .001$, $\eta_p^2 = .635$) even though familiarity scores for the Dutch and Thai drinks were similar for Thai participants, indicating that Thai participants reported to be quite familiar with yogurt drinks. The analysis showed a significant main effect of cultural drink category (*F*(1, 85) = 118.68, $p < .001$, $\eta_p^2 = .583$) on familiarity scores, where the Dutch drink received overall higher familiarity ratings than the Thai drink. In addition, there was a significant main effect of participant nationality (*F*(1, 85) = 57.19, $p < .001$, $\eta_p^2 = .402$), with Thai participants rating significantly higher familiarity scores than Dutch participants.

Explicit measures for universal and cultural food categories (hypotheses 1 and 2)

Universal food category

The mean EmojiGrid valence and arousal scores for all universal food images (regular and molded) rated by both Dutch and Thai participants are plotted in Figure 9-4. The figure shows the typically found U-shaped distribution between valence and arousal scores (Toet et al., 2018a; Kaneko et al., 2019b). As expected, scores for molded food images are situated on the left (low valence) while scores for regular food images are on the right (high valence). The expected culture-dependent response bias was observed in this figure: Thai participants use a smaller portion of the valence and arousal scale compared to Dutch participants. Dutch rated molded food images as more extreme in low valence, and regular food images more extreme in high valence than Thai participants; and Dutch participants rated both images as more arousing compared to Thai participants. These observations are confirmed by the following figures and analyses.

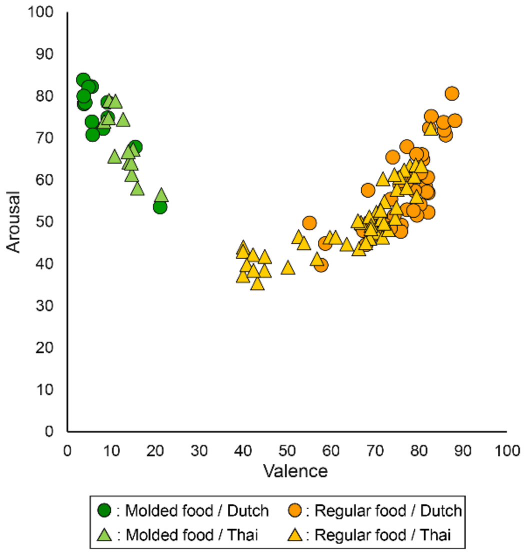


Figure 9-4. Mean valence (x-axis) and arousal (y-axis) scores for all 60 universal (regular and molded) food images. Each data point represents the mean score from either Dutch or Thai participants for each food image.

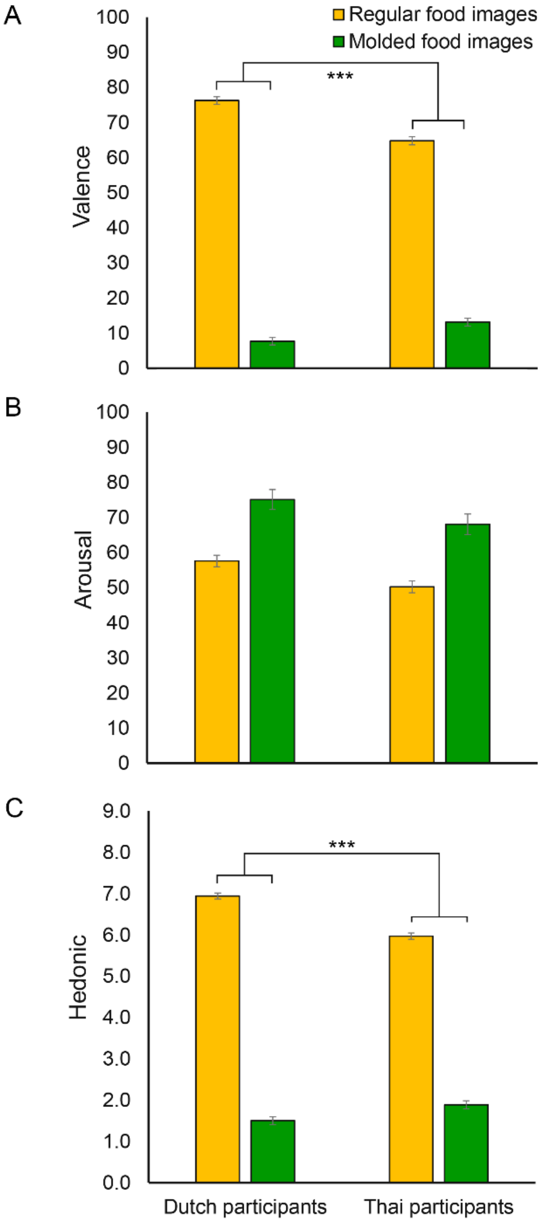


Figure 9-5. The averaged valence (A), arousal (B), and hedonic (C) scores for each participant nationality (Dutch and Thai) and universal food image category (regular and molded). Error bars indicate standard error of the mean. *** indicates a significant interaction between participant nationality and universal food image category with p < 0.001.

Figure 9-5A and B represent average valence and arousal scores for each participant nationality and universal food image category. The two-way mixed ANOVA on valence scores showed a significant interaction between participant nationality and universal food image category ($F(1, 85) = 51.32$, $p < .001$, $\eta_p^2 = .376$), indicating that Dutch participants rated valence more extremely (relatively higher valence rating for regular food images and relatively lower valence rating for molded food images compared to Thai participants). There was also a significant main effect of food image category ($F(1, 85) = 2583.37$, $p < .001$, $\eta_p^2 = .968$), indicating lower rated valence for molded than for regular food images, and a main effect of participant nationality ($F(1, 85) = 9.10$,

$p = .003$, $\eta_p^2 = .097$) with a somewhat higher overall valence score for Dutch than Thai participants. For arousal, there was a significant main effect of participant nationality ($F(1, 85) = 8.89$, $p = .004$, $\eta_p^2 = .095$), indicating higher arousal scores for Dutch participants than for Thai participants. There was also an effect of food image category ($F(1, 85) = 61.33$, $p < .001$, $\eta_p^2 = .419$), showing higher arousal scores for molded food images than for regular food images. There was no significant interaction between participant nationality and universal food image category ($F(1, 85) = .005$, $p = .942$).

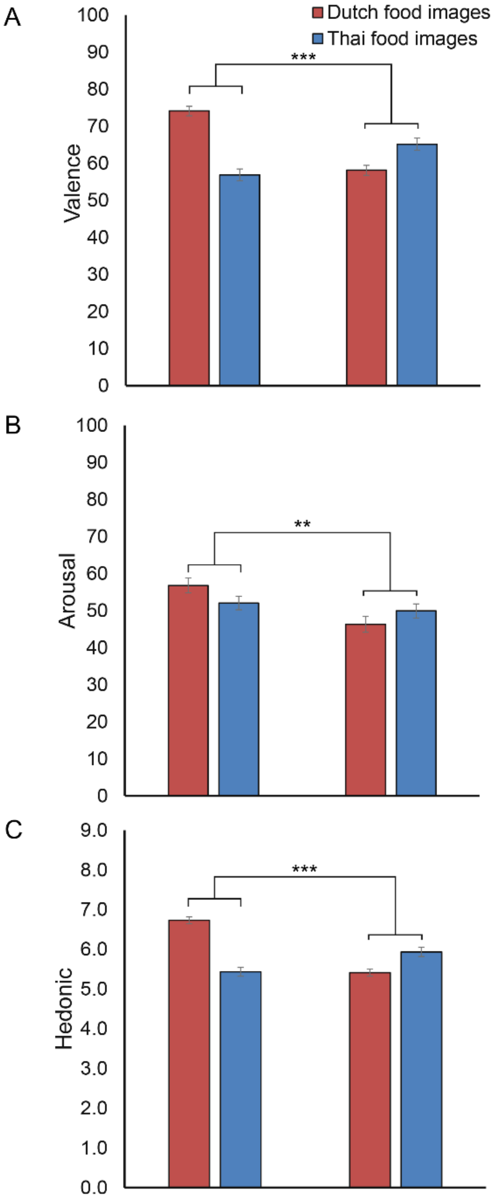


Figure 9-6. The averaged valence (A), arousal (B), and hedonic (C) scores for each participant nationality (Dutch and Thai) and cultural food image category (Dutch and Thai). Error bars indicate standard error of the mean. ** and *** indicate significant interactions between participant nationality and cultural food image category with p < 0.01 and p < 0.001, respectively.

Figure 9-5C shows hedonic liking scores averaged for each participant nationality and universal food image category. The pattern of results was identical to that of valence ratings, with a significant interaction between participant nationality and universal food image category ($F(1, 85) = 63.78$, $p < .001$, $\eta_p^2 = .429$), and main effects of both participant nationality ($F(1, 85) = 11.45$, $p = .001$, $\eta_p^2 = .119$) and universal food image category ($F(1, 85) = 3152.01$, $p < .001$, $\eta_p^2 = .974$).

Cultural food category

The mean EmojiGrid valence and arousal scores for Dutch and Thai cultural food image categories are shown in Figure 9-6A and B. The two-way mixed ANOVA showed the hypothesized interaction between participant nationality and cultural food image category on valence ($F(1, 85) = 129.01$, $p < .001$, $\eta_p^2 = .603$), indicating that Dutch participants rated higher valence for Dutch food images compared to Thai food images, whereas Thai participants showed the opposite pattern. Thus, participants rated food images that match their own nationality as more pleasant. There were also significant main effects of participant nationality ($F(1, 85) = 4.71$, $p = .033$, $\eta_p^2 = .053$) and cultural food image category ($F(1, 85) = 22.82$, $p < .001$, $\eta_p^2 = .212$) on valence scores, indicating overall higher scores for Dutch food images, and higher scores by Dutch participants. As for arousal scores, a significant interaction between participant nationality and cultural food image category ($F(1, 85) = 8.23$, $p = .005$, $\eta_p^2 = .088$) indicated that Dutch participants rated relatively higher arousal for Dutch food images than for Thai food images, while Thai participants showed the opposite pattern. A significant main effect of participant nationality ($F(1, 85) = 7.06$, $p = .009$, $\eta_p^2 = .077$) was found with overall higher arousal scores for Dutch participants. There was no significant main effect of cultural food image category ($F(1, 85) = .15$, $p = .704$).

The mean hedonic liking scores are shown in Figure 9-6C. It shows the same pattern as valence scores, with a significant interaction between participant nationality and cultural food image category ($F(1, 85) = 119.71$, $p < .001$, $\eta_p^2 = .585$), and significant main effects of both participant nationality ($F(1, 85) = 11.98$, $p = .001$, $\eta_p^2 = .124$) and cultural food image category ($F(1, 85) = 21.75$, $p < .001$, $\eta_p^2 = .204$).

Implicit measures for universal and cultural food categories (hypothesis 3 and 4)

Universal food category

Figure 9-7A shows the averaged HR and the HR traces for each participant nationality and universal food image category. As expected, there was a significant main effect of universal food image category on HR ($F(1, 82) = 116.42$, $p < .001$, $\eta_p^2 = .587$), where we observed a lower HR for molded compared to regular images. Also as expected, this effect was the same for participants of both nationalities, indicated by a lack of interaction between participant nationality and universal food image category ($F(1, 82) = .909$, $p = .343$). There was a main effect of participant nationality, ($F(1, 82) = 7.14$, $p = .009$, $\eta_p^2 = .080$) with Thai participants showing higher HR than Dutch participants. Figure 9-7B shows the averaged phasic EDA and the phasic EDA traces for each participant nationality and each universal food image category. The two-way mixed ANOVA on phasic EDA did not show main effects of universal food image category ($F(1, 82) = .166$, $p = .684$) or participant nationality ($F(1, 82) = 2.856$, $p = .095$), and no interaction ($F(1, 82) = .001$, $p = .977$).

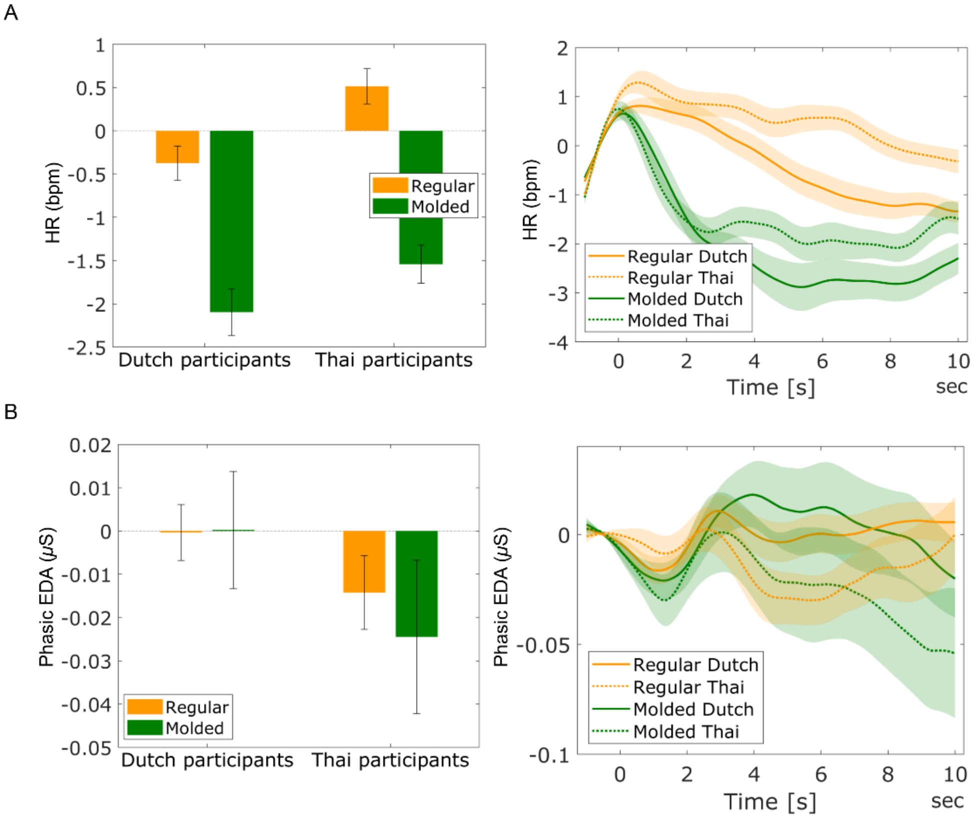


Figure 9-7. The averaged HR (A-left) and phasic EDA (B-left) for each participant nationality (Dutch and Thai) and universal food image category (regular and molded). Error bars indicate standard error of the mean. The figures at the right side show the averaged traces of HR (A) and phasic EDA (B) for each participant nationality and universal food image category during image viewing, with t = 0 indicating fixation cross offset, and food image onset. The light-shaded areas indicate standard error of the mean.

Cultural food category

Figure 9-8A shows the averaged HR for each participant nationality and each cultural food image category. As expected, a significant interaction between participant nationality and cultural food image category was found on HR ($F(1, 82) = 8.50$, $p = .005$, $\eta_p^2 = .094$), with Dutch participants showing lower heart rate for Thai food images compared to Dutch food images, while the pattern was opposite for Thai participants. There was no significant main effect of cultural food image category ($F(1, 82) = 2.93$, $p = .091$), but there was a significant main effect of participant nationality ($F(1, 82) = 10.05$, $p = .002$, $\eta_p^2 = .109$) with Thai participants showing higher HR than Dutch participants. Figure 9-8B shows the averaged phasic EDA and the phasic EDA traces for each participant nationality and each cultural food image category. For phasic EDA, no interaction between nationality and cultural food image category was found ($F(1, 82) = 3.02$, $p = .086$), and no significant main effect of cultural food image category $F(1, 82) = 2.74$, $p = .101$). There was a significant main effect of participant nationality ($F(1, 82) = 4.28$, $p = .042$, $\eta_p^2 = .050$) with Dutch participants showing higher EDA than Thai participants.

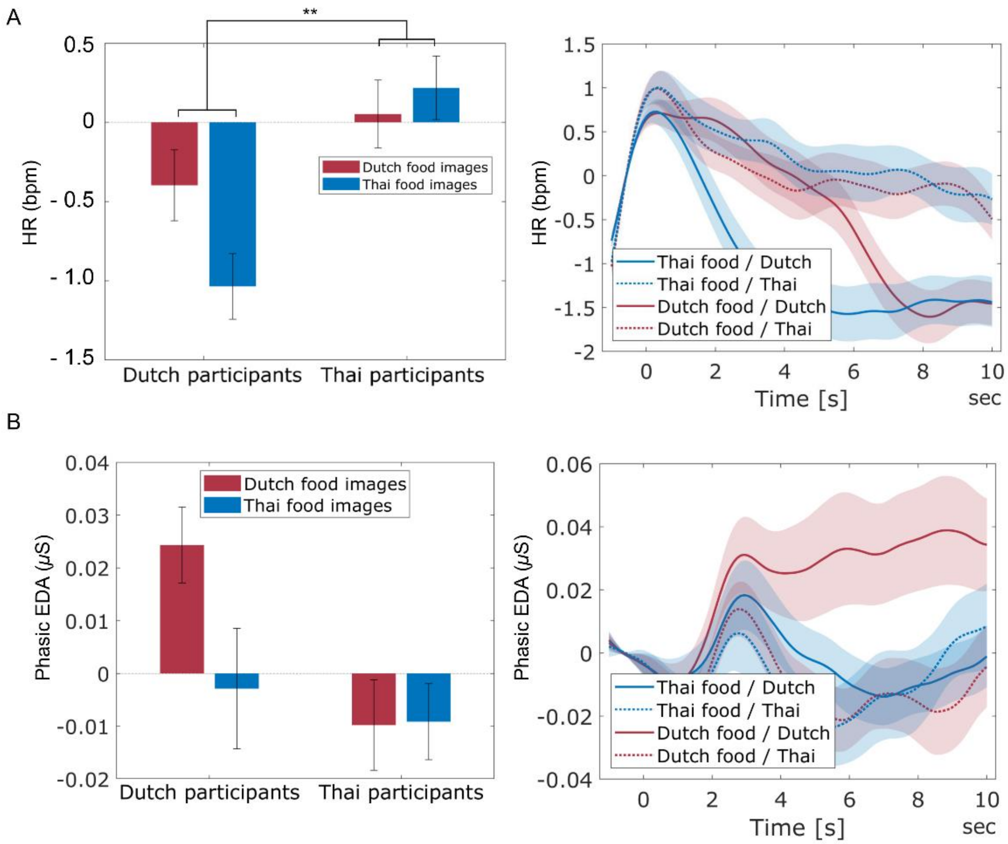


Figure 9-8. The averaged HR (A-left) and phasic EDA (B-left) across participant nationality (Dutch and Thai) and cultural food image category (Dutch and Thai). Error bars indicate standard error of the mean. The figures at the right side indicate the averaged trace of HR (A) and phasic EDA (B) across participant nationality and cultural food image category during image viewing, with t = 0 indicating fixation cross offset, and food image onset. The light-shaded areas indicate standard error of the mean. ** indicates a significant interaction between participant nationality and cultural food image category with p < 0.01.

Explicit and implicit measures for cultural drinks (hypothesis 5)

Explicit measures

Figure 9-9A, B, and C show respectively the average valence, arousal, and hedonic liking ratings for each participant nationality and each cultural drink. The two-way mixed ANOVAs did not show any interaction or main effects of cultural drink category and participant nationality on any of the three measures (all *p*-values are .094 or higher).

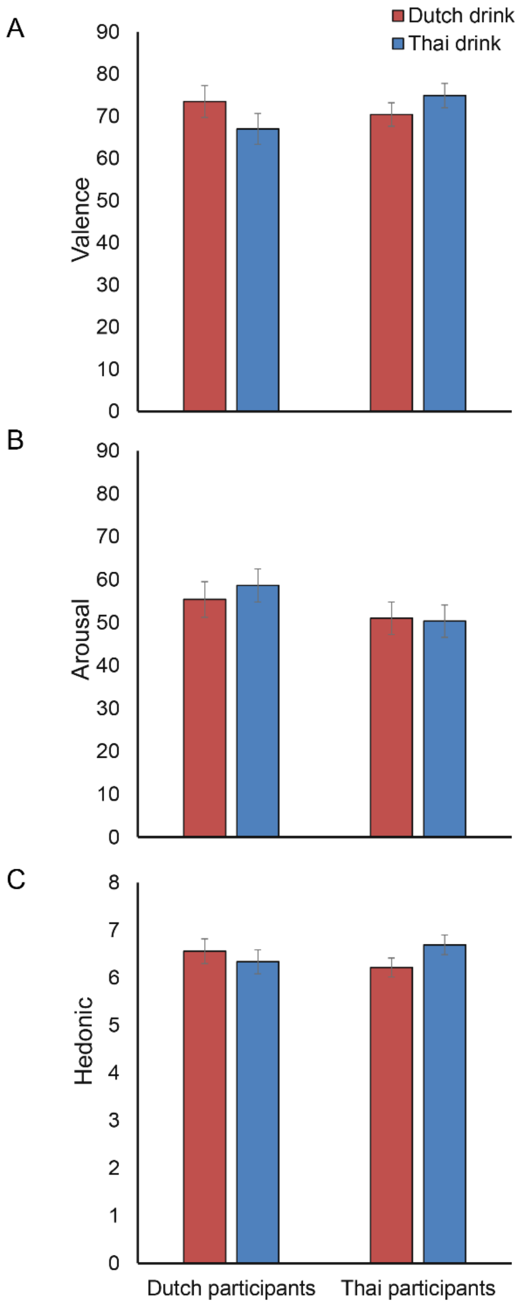


Figure 9-9. The averaged valence (A), arousal (B), and hedonic (C) scores across participant nationality (Dutch and Thai) and cultural drink category (Dutch and Thai). Error bars indicate standard error of the mean.

Implicit measures

The average HR and phasic EDA for each participant nationality and each cultural drink are shown in Figure 9-10A and B. The two-way mixed ANOVA showed the hypothesized interaction between participant nationality and cultural drink category on HR ($F(1, 82) = 5.79$, $p = .018$, $\eta_p^2 = .066$). While both nationalities tend to show a higher HR for the Dutch drink, this is especially the case for Thai participants. Participant nationality ($F(1, 82) = 5.72$, $p = .019$, $\eta_p^2 = .065$) and cultural drink category ($F(1, 82) = 10.46$, $p = .002$, $\eta_p^2 = .113$) showed main effects on HR, with overall higher HR for the Dutch drink than the Thai drink, and higher HR for Thai participants than Dutch participants. For phasic EDA, there was no interaction between participant nationality and cultural drink category (*F*(1, 82) = 1.07, *p* = .303), and no significant main effect of cultural drink category (*F*(1, 82) = 2.96, *p* = .089). Dutch participants showed overall higher phasic EDA than Thai participants (main effect of participant nationality: *F*(1, 82) = 5.62, *p* = .020, η_p^2 = .064).

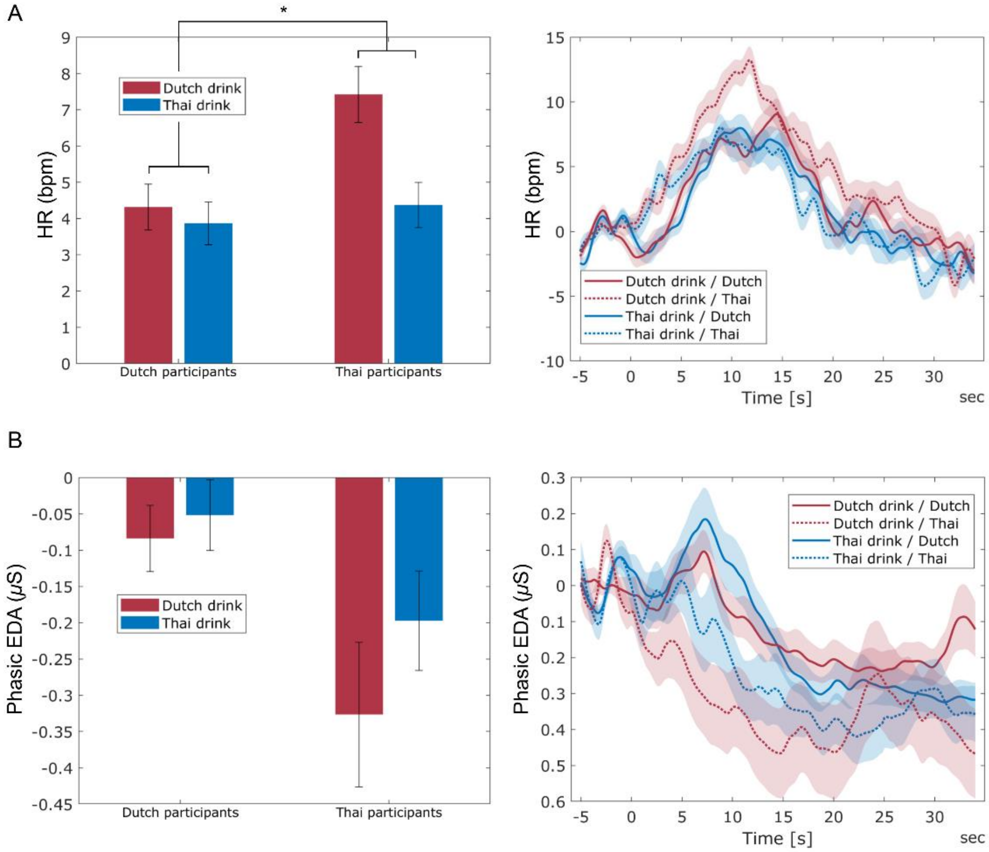


Figure 9-10. The averaged HR (A-left) and phasic EDA (B-left) across participant nationality (Dutch and Thai) and cultural drink category (Dutch and Thai). Error bars indicate standard error of the mean. The figures at the right side indicate the averaged trace of HR (A) and phasic EDA (B) across participant nationality and cultural drink category. The name of the drink is presented from t = −5 to t = 0, after which participants took one sip. The light-shaded areas indicate standard error of the mean. * indicates a significant interaction between participant nationality and cultural drink category with $p < 0.05$.

Figure 9-11 shows the averaged behavioral measure of sip size. The expected interaction between participant nationality and cultural drink category was found (*F*(1, 85) = 15.24, *p* < .001, η_p^2 = .152), indicating that Dutch participants took a larger sip of the Dutch drink than the Thai drink, while this was opposite for Thai participants. A significant main effect of participant nationality was found (*F*(1, 85) = 6.97, *p* = .010, η_p^2 = .076) indicating that Thai participants took larger sips in general. There was no main effect of cultural drink category (*F*(1, 85) = .82, *p* = .367).

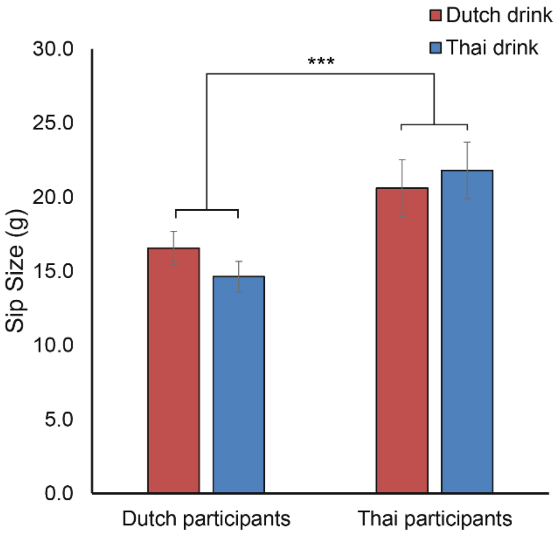


Figure 9-11. The averaged valence (A), arousal (B), and hedonic (C) scores across participant nationality (Dutch and Thai) and cultural drink category (Dutch and Thai). Error bars indicate standard error of the mean.

Discussion

The present study investigated the potential of implicit physiological measures to provide objective measures of affective food experience in contrast to explicit self-report measures, for Dutch and Thai participants. Explicit self-reports of these participants are expected to be influenced by differential cultural bias, therewith hampering comparison of food experience across cultures. Implicit physiological measures could potentially solve this problem.

Explicit self-report responses towards universal food image categories (regular and molded food images) revealed the usage of an extreme response style by Dutch participants compared to Thai participants who used a middle response style, which is consistent with the literature on response style characteristics of cultural groups (Chen et al., 1995; Johnson et al., 2005; Kaneko et al., 2019c), and confirming hypothesis 1. The valence and liking scores were higher at both ends of the scale for Dutch participants than for Thai participants. While there was no such extreme response style at both ends of the scale for arousal (where in contrast to valence, no obvious neutral location exists), Dutch participants rated universal food images more extremely on high arousal compared to Thai participants.

Familiarity ratings of the cultural food images confirmed that for Thai participants, Thai food images were more familiar than Dutch food images, while the reverse was true for Dutch participants. This confirmed that our food image stimuli were properly selected. As hypothesized in hypothesis 2, significant interactions between participant nationality and cultural food image category for self-reported valence, arousal, and hedonic liking revealed that participants rated food images from their own cultural food images as more pleasant and arousing than from the other culture. Higher liking scores for familiar foods is consistent with previous studies (Torrico et al., 2019).

As stated in hypothesis 3, HR responses were sensitive to affective food experience, as indicated by an effect of universal image type (lower HR for molded compared to regular images). The direction of this effect is consistent with literature on heart rate responses to high and low valence images (Bradley et al., 1990; Lang et al., 1993). As also stated in hypothesis 3, this effect was the same for both cultural groups, suggesting that despite the culturally-dependent difference in explicit ratings, affective food experience of universal food images (regular and molded) is the same across cultures. For cultural food images, we found the expected interaction effect between participant nationality and cultural food category on HR (hypothesis 4). Dutch participants had a lower HR in response to Thai food images compared to Dutch food images and *vice versa* for Thai participants. The direction of the effect is as expected, consistent with lower valence for unfamiliar food.

For cultural drinks, we found an interaction between participant nationality and cultural drink category on the implicit behavioral measure of sip size in the expected direction (hypothesis 5), indicating that participants of both nationalities take a larger sip of the drink from their own culture. HR also showed an interaction between participant nationality and cultural food category, where Thai participants showed a particularly strong response to the Dutch drink. These results corroborate previous results on sip size and HR (and other physiological measures) as sources of information on affective experience of taking a sip of a drink (Kaneko et al., 2019a), where sip size was smaller and HR was higher for (low valence and) high arousal drinks. In the present study, visual, olfactory and taste properties of the drinks were held constant and only the associated emotion differed. In contrast to the sip size and HR results, we unexpectedly did not find an interaction effect between participant nationality and cultural food category in explicit self-reported affective scores. Also, while familiarity scores showed the intended interaction between participant nationality and cultural food category, and Dutch participants rated the Dutch drink as much more familiar than the Thai drink, Thai participants rated the two drinks almost equal in familiarity, which is in apparent conflict with their smaller sip size and higher HR response. Note that in contrast to the implicit and explicit responses to food images, in the case of taking a sip, participants' sensory perception fundamentally changes over time. It starts with an expectation primed by the announcement of the drink name, is followed by perceiving its visual and olfactory properties, and ends with the actual taste. We speculate that the seemingly conflicting results between implicit and explicit measures are caused by their difference in time. For measures that reflect the time of taking a sip and shortly thereafter (sip size and HR), we observe a pattern in line with the hypothesis. When rating the drinks, which happens about half a minute later, Thai participants may realize they are actually familiar with the yogurt drink, and participants of both nationalities may like the taste of both drinks.

In contrast to HR, phasic EDA was not a sensitive measure of food evoked emotion in the present study, as indicated by our finding that phasic EDA responses did not differ between regular and molded food images. In addition, no interaction between participant nationality and cultural food category was found for phasic EDA; neither for images nor for drinks. While in general, there is no straightforward relation between HR and valence, studies using emotional images as stimuli consistently show valence (rather than arousal) effects, where pleasant stimuli correlate with higher heart rate acceleration than unpleasant stimuli (Hare et al., 1970; Libby Jr et al., 1973; Winton et al., 1984; Greenwald et al., 1989; Bradley et al., 1990; Lang et al., 1993; Lang et al., 1998; Bradley and Lang, 2000; Anttonen and Surakka, 2005; Codispoti and De Cesarei, 2007; Sokhadze, 2007). We found this HR effect as well, both in universal and cultural food categories. For EDA, consistent positive associations with arousal have been found in a range of contexts, including emotional/neutral picture viewing tasks (Greenwald et al., 1989; Lang et al., 1993; Lang et al., 1998) and tasting drinks (Kaneko et al., 2019a; Brouwer et al., 2020). In the present study, phasic EDA showed no (interaction with) food category effects. We did not have a clear a priori expectation that arousal should differ between cultural food image categories (familiar and unfamiliar), but we had expected higher arousal for molded food images compared to regular food images. This expectation was supported by the ratings, but the difference in rated arousal between regular and molded food images was modest compared to the difference in valence. Thus, the lack of effect in EDA may have been caused by relatively small genuine difference in arousal between our stimulus categories; they differed more in valence which, at least for images, is better captured by HR. Consistent with our findings is a study by Anderson et al. (2019). This study showed similar skin conductance levels when viewing both rotten and sweet food images, while HR was decreased when viewing molded food images compared to sweet food images.

Our analyses showed trends indicating overall higher HR and lower phasic EDA for Thai participants than for Dutch participants. Note that our HR and phasic EDA variables reflect baselined responses. We examined whether these differences in responses to food stimuli might be related to differences in overall HR and phasic EDA, which e.g. could have been caused by a difference in climate between the Netherlands and Thailand (Madaniyazi et al., 2016). However, we did not find statistically significant differences between nationalities for unbaselined HR and EDA levels.

While implicit physiological measures have been praised as objective markers of affective experience, we do not know of studies that clearly show or suggest that explicit self-reported ratings are less accurate markers of food experience than physiological measures. With our study, we attempted to fill this gap, and our results suggest that implicit physiological measures could indeed be considered as better indicators of food experience than explicit ratings. However, we should consider that we do not (cannot) know the absolute ‘true’ emotion. Our findings could still be consistent with a genuine difference between Thai and Dutch participants with respect to the experience of viewing regular and molded food images, following the explicit ratings which are not ‘really’ biased, but reflecting genuine differences. In this interpretation, HR would not be sensitive enough to capture this, even though HR could capture other effects that may have been stronger. Replication of our results with other types of ground truth stimuli, and relating emotions to behavior (e.g. food choice), would therefore be good to further establish the findings of the present study. For future studies, it would also be of interest to investigate if and how possible cross-cultural differences in food neophobia influence different measures of affective food experience. Finally, for adding another dimension of food experience, it would be valuable to include implicit measures of approach-avoidance motivation such as EEG alpha asymmetry

(Harmon-Jones et al., 2010; Brouwer et al., 2017d) or a behavioral approach-avoidance task (Solarz, 1960; Piqueras-Fiszman et al., 2014; van Beers et al., 2020).

Conclusion

In the present study, we examined implicit measures of food experience in a case that self-reported ratings cannot be taken at face value because of possible culturally dependent response bias. Our study confirmed the existence of such a cultural response bias and showed that only the implicit physiological measure of HR followed the prior expectations of genuine food experience. Different contexts, in this case, different types of food stimuli (images and drinks) resulted in different sensory, affective and dynamics of physiological responses. For both types of stimuli, we found the expected interaction between participant nationality and cultural food category on implicit measures (HR, and in case of tasting, sip size). This study indicates that physiological responses can be used to investigate differences in affective food experience between cultures. Especially when estimating possible acceptance of products and product promotion in cultures that strongly differ with respect to expressing affect, such as Asian and western cultures, using self-reports alone may lead to incorrect conclusions.

Acknowledgements

We thank Prof. Suwimon Keeratipibul, Dr. Punnida Techaruvichit, and Dr. Panita Ngamchuachit from Chulalongkorn University in Bangkok, Thailand, for their supports of recruiting Thai participants and providing a useful testing ground for the experiment conducted in Thailand. Also, we thank Dr. Nattapong Thammasan from University of Twente in Enschede, the Netherlands, for his assistance with conducting the experiment in Thailand and thank Mr. Jan Ubbo van Baardewijk for extracting and analyzing physiological data. We also acknowledge the people of ‘The Streetfood Bar’ in Amersfoort, the Netherlands, for their assistance in creating Thai food images.

Chapter 10

General Discussion

This thesis investigated what (type of) measures are used to evaluate food-evoked emotions and to what extent these measures are able to capture food-evoked emotions. Also, we explored the added value of the different measures and their combination, all of these specifically in a cross-cultural context, where the same foods can be either novel or familiar, and where measures other than explicit, verbal measures are expected to be especially valuable. A new cross-cultural food image database and pictorial emotional response tool were developed as well. This thesis contributed to the methodology to more comprehensively and reliably measure food related emotions in cross-cultural contexts. Below general conclusions are drawn and outstanding issues identified and discussed.

Measures to estimate food-evoked emotions

The literature review in this thesis showed that previous studies predominantly rely on explicit, verbal ratings and questionnaires or a combination of redundant explicit measures to assess food-evoked emotions, despite the drawbacks of explicit measures as described in the Introduction of this thesis. More than 80% of the 101 papers selected for the literature review used hedonic scaling measures to evaluate the degree of food-evoked emotions (see Figure 2-4). A limited number of studies use implicit physiological and behavioral measures to evaluate emotion elicited by food experience. The 3 x 3 complementary toolbox table (shown in Table 2-2), that classifies measurement tools along three emotional processing levels and three methodological levels, may motivate researchers to use a complementary combination of physiological, behavioral, and cognitive measures to cover all three emotional processing levels instead of relying only on self-report questionnaires. In line with this, de Wijk and Noldus (2021) argue in their recent review that it is important to integrate responses from both conscious and unconscious processing obtained by explicit and implicit measures, respectively, to understand emotion evoked by food experiences. Given the argument that food-evoked emotion and food choice are related (Dalenberg et al., 2014; Gutjar et al., 2015b), choosing more appropriate measures or proper combinations of both explicit and implicit measures may be a definite step to better predict an individual's actual food choice and buying behavior.

As described in the literature review, explicit self-report questionnaires are the most widely used tools to evaluate food-evoked emotions since many of these have been extensively validated and used before, and are easy to apply for many researchers. In self-report questionnaires, participants' affective state and food evoked emotions are usually evaluated by asking participants to select appropriate emotional words and to rate the degree to which each of them applies. However, it can be difficult to verbally describe food evoked emotions, the emotional lexicon varies across cultures and is often not easily translated between different languages. Besides using measures that are not cognitive in nature, a solution to deal with the translation challenge when measuring cross-culturally is the pictorial self-report tool (EmojiGrid).

Previous studies provide some evidence that Emojis have the potential to overcome the difficulty to translate the emotional lexicon in one language to another (Jaeger et al., 2017b; Schouteten et al., 2018). In this thesis we combined Emojis with the valence-arousal affect grid (Russell and Barrett, 1999) and developed an intuitive, language-independent graphical tool called "EmojiGrid" (Chapter 7, 8). The study in Chapter 7 showed that each of the 16 Emojis used in the grid conveyed the intended degree of valence and arousal, apparent in the high correlation between these Emojis and a conventional graphical tool called Self-Assessment Manikin (SAM) (Bradley and Lang, 1994). Also, this chapter demonstrated that the EmojiGrid is able to capture the affective dimensions of valence and arousal for food image stimuli. Chapter 8 further explored the potential of "EmojiGrid" for cross-cultural studies with respondents with different mother languages (three European countries (Germany, the Netherlands, and United Kingdom) and one Eastern country (Japan)) through an online experiment with 60 universal food images covering a wide range of valence (from highly unpleasant to very pleasant) obtained from a CROss-CUltural Food Image Database (CROCUFID). The study showed that EmojiGrid reproduces the nomothetic U-shaped relation (Chapter 7, 8 and also see (Toet et al., 2019b) for similar findings for odor-evoked emotions) between valence and arousal ratings among the four different countries. Another important result obtained in this study is that valence and arousal ratings with EmojiGrid were stable in three experiments conducted in Chapter 7, while (arousal) ratings through conventional verbal questions with VAS scale depended on the exact formulation/lay-out of the questions. The EmojiGrid has now been successfully validated in various other studies, e.g. on odors (Toet et al., 2019a) and experienced emotions and perceived emotions evoked by images of objects and landscapes and black-and-white line drawings of people (Toet and van Erp, 2019). It was also used in a study on arousal evoked by tasting white rice with familiar and unfamiliar soy sauce (de Wijk et al., 2021). This study showed that the EmojiGrid is more sensitive to arousal in response to food's intrinsic taste compared to the extrinsic food cues, such as brand information. In sum, the Emojigrid appears a valuable new tool, that while it is an explicit tool, it is independent of language.

The research findings described in this thesis indicate the potential of implicit measures for capturing food-evoked emotions (Chapter 3, 4, 5, and 9). To the best of our knowledge, the study described in Chapter 3 is the first study to simultaneously examine the sensitivities of several different explicit and implicit measures for monitoring the emotions evoked by tasting drinks (though a few studies exist that used explicit and implicit measures to assess emotion elicited by food samples). As expected, explicit valence and arousal ratings showed high discriminative power between the regular drinks and the vinegar solution, and an ANOVA testing the effect of the regular drinks on explicit self-report ratings suggested that participants also agreed on small differences in affective experience. One important finding here is that besides explicit self-report ratings, implicit physiological and behavioral measures used in this study (heart rate (HR), electrodermal activity, heart rate variability (HRV), pupil size, facial expression, frontal alpha asymmetry (FAA) from EEG, and sip size) are also sensitive enough to discriminate food evoked emotions at least between samples that differ strongly in terms of elicited emotion (regular drinks and a vinegar solution).

Among the implicit measures evaluated in this thesis, sip size had the largest power to discriminate the vinegar solution from the regular drinks (Chapter 3). The thesis further showed that the sensitivity of sip size was robust at least one week later (Chapter 5). Sip size also followed the prior expectations of genuine food experience in a cross-cultural study, indicating that both Thai and Dutch participants took a larger sip of the drink from their own culture (Chapter 9). Given that food and drinks from the own culture are preferred, this is in line with previous studies that showed that the amount of food intake is associated with food liking (Zandstra et al., 2000).

Besides the implicit behavioral measure of sip size, HR (with the advantage of being both implicit and continuous) appeared to be sensitive to large as well as subtle differences in affective experience. Of the implicit measures, heart rate most stably reproduces the effects throughout the studies (Chapter 3, 4 and 9). In the study described in Chapter 3, HR strongly increased when tasting an a-priori known low-valence and high-arousal vinegar solution compared to the regular drinks. In addition, HR had significant positive correlation with arousal ratings among the regular drinks. Also, the study described in Chapter 4 showed that HR was suitable to train a model that predicted arousal. Furthermore, HR responses reflecting arousal were stable across cultures (the cross-cultural study between Thai and Dutch participants described in Chapter 9). It should be noted that our result that HR increases with arousal does not hold across all types of affective studies, e.g. studies using IAPS non-food emotional pictures as stimuli or reading emotional sections in a book (Brouwer et al., 2015b) where opposite relations have been found. As another example, Walsh et al. (2017) showed that HR decreased after participants saw spoilage quality concern evented video stimuli while HR increased when facing hygiene quality concern evented video stimuli, despite that both video stimuli may be considered as high arousal, negative emotions. Therefore, although HR can be considered as a good implicit indicator to evaluate food-evoked emotion in the contexts tested by us, it is important to include a ground truth emotional stimulus (i.e., a stimulus of which there is no doubt concerning the emotional response of most participants) in other contexts to check if and how HR responds in that situation.

Besides HR and sip size measures, other implicit measures such as skin conductance, facial expression, pupil dilation, HRV and FAA from EEG were also evaluated (Chapter 3). This study indicated that skin conductance, facial expression, pupil size, and FAA showed discriminative power for cases in which food experiences differ strongly while they appeared not to be sensitive enough to detect subtle emotional differences among the regular drinks. On the other hand, HRV did show a significant correlation with explicit ratings of arousal without considering the vinegar solution, and this may help to increase the validity and reliability of rank ordering regular drinks with subtle emotional differences along arousal scales besides HR and sip size.

Our cross cultural study showed that phasic electrodermal activity (phasic EDA) responses were similar even between regular and molded food images (used as extreme stimuli) (Chapter 9), which is consistent with the findings reported by Anderson et al. (2019) that similar skin conductance level is obtained when viewing both rotten and sweet food images. On the other hand, previous studies suggest that skin conductance results are more stable and more reliable than heart rate in the context of food expectation (Tena, 2019). A challenge of physiological implicit measures is that physiological responses reflect a mixture of factors besides emotional response to food stimuli, including other mental states such as attention, as well as body movements. Related to this, the results of implicit measures may depend on the specific stimuli and type and phase of food experience one intends to evaluate. Also, the way that physiological signals are recorded and analyzed widely differs across studies and can impact results. Researchers should carefully consider the context and the purpose of the research when conducting an experiment with implicit measures, and as mentioned before, include a ground truth emotional stimulus to support interpreting results

Measures of emotions evoked by experiencing novel (other culture) and familiar (own culture) foods

The results in this thesis underscore and extend the importance of food familiarity on food experience that has been found in the literature before (e.g. Fenko et al. (2015b)), where food familiarity depends on an individual's (food) cultural background. Using both explicit and implicit measures, we evaluated the effect of emotional state (both positive and negative) and food familiarity on food-evoked emotions (Chapter 5). This study showed that a familiar soup was rated more positively than a novel soup, especially in the negative emotional condition, confirmed by both explicit (valence rating and EsSense 25 emotional questionnaire) and implicit measures (sip size and willingness-to-take-home). Both measures also provided important insights that the effect of emotional state was robust over at least one week. The cross-cultural study in Chapter 6 also showed that participants rated own regional food images higher in valence and arousal compared to unfamiliar food images. While these explicit measures reproducibly reflect similar emotions to novel and familiar foods, it should be noted that explicit ratings can be misleading due to response bias across cultures, such as different response styles: a Middle Response Style (MRS) is often obtained from Asian participants, and an Extreme Response Style (ERS) is often seen in Western regions (Harzing, 2006). Also, a direct comparison of affective food ratings across cultures is typically compromised by the differential influence and interpretation of verbal scale anchors (Ares, 2018). The cross cultural study described in Chapter 9 demonstrates that implicit measures (especially HR) are not affected by response bias across cultures while explicit measures are, implying that implicit measures can be beneficial in cross-cultural comparisons of food-evoked emotions. Specifically, this study showed the expected response bias for Thai and Dutch participants in explicit ratings while implicit measures of heart rate (in case of food images) and sip size (in case of drinks) followed the prior expectations of genuine food experience, which is consistent to our results in Chapter 3 and 4.

Despite the importance of novel and familiar food images as stimuli in research on affective and appetitive responses across cultures (e.g., in online experiments), there are no publicly available food image databases that cover the affective space of valence and arousal, include a wide range of foods and cuisines from all over the world, and are developed with the use of standardized recording methods. To facilitate cross-cultural studies on food-evoked emotions, this thesis provided a new food image database (CROCUFID) compensating for those shortcomings and reports the first cross-cultural study to validate it in an internet survey (Chapter 6). It now contains more than 980 food and non-food images, covering Asian and Western cuisines, and the content is still growing through contributions of researchers from different countries. Tahir et al. (2021) newly proposed a food-related emotion classification model using CROCUFID. We expect this database to help sensory and consumer researchers and food industries to conduct studies across cultures.

Combining measures

We evaluated the sensitivity of several explicit and implicit measures to capture food-evoked emotions. For implicit measures, we showed that most are sensitive to emotions to some degree, specifically to the arousal dimension of the emotion. Previous studies have used explicit ratings of hedonic liking, valence (pleasantness), and willingness to try while measuring arousal has been neglected. A reason for this could be that it is a relatively difficult concept to grasp and self-report on, especially for non-native English speakers because no appropriate word for arousal exists in many other languages. However, if one is interested in assessing food products, which are usually all about equally pleasant, arousal may be of special interest since the level of arousal affects the memorability of an event and is considered to be a crucial determinant in defining (sustained) attractiveness of products (Bradley et al., 1992; McGaugh, 2006). Köster and Mojet (2015b) also argued that arousal level is a crucial factor to assess attractiveness of food products in food marketing because levels of arousal are related to the memorability of an event. Implicit and explicit measures seem to complement each other nicely with respect to valence and arousal.

This thesis also explored the potential of combining implicit behavioral and (neuro)physiological measures to estimate affective food experience when tasting drinks without relying on explicit measures at all (Chapter 4). Two models were trained using different approaches: one model was trained using a traditional approach that considered the explicit valence and arousal ratings as the ‘true’ affective experience of drinks, and the other model was trained by using the vinegar solution as a ground truth of extreme low valence and high arousal; and regular drinks as high valence and low arousal. While the conventional approach of using explicit valence and arousal ratings as ground truth failed to provide a sophisticated predicting model, the predicting model trained by using the vinegar solution as ground truth (low-valence and high-arousal) resulted in a high correspondence between estimated arousal by the model and self-reported arousal among subtle different regular drinks. Also, this study showed that the best combination of any of two implicit measures is heart rate and sip size, which corresponds to the outcomes of the study described in

Chapter 3. This study indicated that it is possible to estimate emotional experience elicited by tasting a regular drink using only implicit data without any explicit self-reports.

An advantage of many implicit physiological measures that complement explicit measures, is their continuous nature – they could e.g. provide a pattern of arousal over the entire food experience. Generally speaking, food-evoked emotions are often evaluated “during” food experience, not “before” and “after” although in a real life such emotions also happen before and after the actual tasting and are dynamic rather than static (see Chapter 3 and 5, also see (Sander et al., 2005; Jager et al., 2014). As argued by Köster and Mojet (2015b), food-evoked emotions should be evaluated not only during sensing/tasting food, but also before tasting (expectation) and after tasting (memory of food experience). This thesis provided the temporal patterns of physiological responses (e.g. heart rate, skin conductance, pupil size, and EEG) that turn out to follow a certain dynamic pattern during an entire tasting consisting of reading the name of a sample, taking one sip, and sitting still for a certain period of time after tasting (in case of the study in Chapter 3), which may be difficult to rate with explicit self-report measures. While we did not attempt to evaluate to what extent these dynamic effects exactly correspond to explicitly reported experience over this short time span, a study that did examine the continuous use of implicit, physiological measures for evaluating food experience across a longer time span, namely during cooking (Brouwer et al., 2019), found some evidence for its validity.

Meiselman (2021) argues that there is not one measure that completely captures food-evoked emotions, instead it is important to find a best combination of measures. In line with his argument, this thesis proposes that understanding and measuring food-evoked emotions in an entire food experience with the 3 x 3 tool box table as a reference should be conducted as the next step in the future of food related consumer science. Among several implicit physiological and behavioural measures, heart rate and sip size (amount of intake) may be an ideal combination of measures to evaluate food-evoked emotions besides self-report questionnaires. We emphasize that in some cases, it may not be useful to add implicit measures, but it is or can be for cases where one expects a strong bias in explicit measures (including response bias across cultures, or biases induced by social pressure), and cases where a continuous measure of food-evoked emotions is desired. Generally, one should consider that explicitly asking to rate food-related emotions will interfere with the actual food experience itself (Winkielman et al., 2011b).

Some important aspects from a commercial point of view

Due to globalization in the food market, it is a critical issue for international food industries and marketeers to understand how consumers with different food cultural backgrounds emotionally respond to food products. It is easily understandable that people who grew up in an Asian food culture will differently respond to a food product compared to those raised in Western countries. At this moment, many food industries and marketeers rely solely on consumers’ self-report questionnaires to understand their response on food products although there are several drawbacks as described in this thesis, such as culturally determined explicit response biases. Our studies aim to investigate to what extent implicit measures are applicable to assess consumers' food-evoked emotions, and the overall findings show that implicit physiological and behavioural measures have definite potential to complement self-report questionnaires. Tena (2019) argues that capturing autonomic nervous system responses (heart rate and skin conductance), representative of implicit physiological responses, may not serve any practical purposes (e.g. food marketing). However, we think that these implicit measures help to better understand the levels of arousal induced by food experience besides the valence ratings measured by conventional self-report questionnaires. Prescott (2017) argues in a review that emotions expressed as levels of valence and arousal may be powerful predictors of consumers' food behaviour, and that arousal is key to reinforce the memory of a certain food product and food experience, which connects to an important marketing strategy for food industries to have more repeating customers in a market.

A limitation of implicit measures in terms of a commercial point of view in a cross-cultural context is that recordings are presently made in a laboratory, resulting in a low number of participants and requiring that visits of each country to recruit local participants for cross-cultural studies. Also, evaluating food-evoked emotions in a lab is only a first step in predicting consumers' food choice and behaviour in real life situation. Walsh et al. (2017) stress that food experience when wearing devices such as EEG caps and HR and skin conductance electrodes, differs from food experience in participants' daily life situation. Development of devices measuring physiological responses remotely will alleviate these problems and facilitate researchers and food industries to capture consumers' food-evoked implicit emotions in their real life situation. An example of recent developments in this area is remote HR detection based on facial skin color changes which allows researchers to evaluate heart rate from video images of the face (remote photoplethysmography or RPPG) (Wei et al., 2012; Wang et al., 2016b) and may enable capturing participants' food-evoked emotions on a daily basis.

Conclusion and future studies

In this thesis, explicit and implicit measures that are used to evaluate food-evoked emotions were reviewed, and we evaluated the potential of several different implicit measures as well as a combination of those measures to better understand food-evoked emotions. We showed that explicit measures are much more often used compared to implicit measures and recommend using multiple measures with different emotional processing levels from the 3 x 3 toolbox table. Among implicit measures evaluated in this thesis, HR and sip size were stable and reproducible measures even in a cross-cultural context. Furthermore, we developed a new intuitive language-indepenent affective evaluation tool, EmojiGrid, and a new open-source standardized food image database, CROCUFID, covering not only Western cuisine but also Asian food to facilitate cross-cultural studies in the food domain.

A recent development in implicit behavioral measures relates to the approach avoidance task (AAT). The principle governing the approach-avoidance is that the response behavior of an individual towards a given stimulus is spontaneously modulated by their perceived valence of the stimulus; there is an approach motivation towards positively evaluated, attractive stimuli (positive valence) and an avoidance motivation away from negatively appraised, aversive stimuli and events (negative valence) (Cacioppo et al., 1993). Previous studies used this approach avoidance tendency to evaluate consumers' implicit-subconscious responses towards food stimuli with joystick (Piqueras-Fiszman et al., 2014; Kytö et al., 2019) or with touchscreen-based arm movements (Meule et al., 2019; Meule et al., 2020). Recently, Zech et al. (2020) newly developed a mobile version of AAT (mAAT). We found that it can be used as a new implicit behavioral measure to assess arousal evoked by viewing food images (Brouwer et al., 2021). It is eagerly anticipated that such new intuitive assessing tools and implicit behavioral measures are further developed and validated.

Implicit measures do not map one-on-one to single emotions. One should always keep in mind that implicit measures reflect other processes besides emotions and depend on aspects such as the stimulus and the analysis method used. Future research using measuring devices that do not disrupt consumers' food experience and can be used remotely would facilitate capturing genuine food-evoked emotions in real life, and enable researchers to investigate what emotions and/or other factors would predict repeated choosing/buying behavior.

Summary

Traditionally, consumer science groups and food companies use self-report measures, in which people indicate how much they like a certain food after having tasted it, to predict consumers' future buying behaviors. However, it has been suggested that food-evoked emotion may improve prediction of individuals' food choice. It may reveal previously unknown product attributes that go beyond traditional sensory and acceptability measurements, and can be a valuable source of information for product development and marketing. Although many research groups acknowledge that the evaluation of food-evoked emotions is probably important to predict food choice behavior, and although many different emotion-association questionnaires and measures have been developed and used, it is still unclear how to best assess food-evoked emotions.

This thesis investigated what (type of) measures are used to evaluate food-evoked emotions and to what extent these measures are able to capture food-evoked emotions. Also, we explored the added value of the different measures and their combination, all of these specifically in a cross-cultural context, where the same foods can be either novel or familiar, and where measures other than explicit, verbal measures are expected to be especially valuable. A new cross-cultural food image database and pictorial emotional response tool were developed. This thesis contributed to the methodology to more comprehensively and reliably measure food related emotions in cross-cultural contexts. To achieve the goals, this thesis consists of two sections, 'Measures to evaluate food-evoked emotion' and 'Measures of food-evoked emotion in cross-cultural context'.

In the first part of the thesis, the following topics are addressed; what (type of) measures are currently used to evaluate food-evoked emotion (**Chapter 2**), to what extent explicit ('conscious') and implicit ('unconscious') measures are able to capture food-evoked emotion by tasting drinks (**Chapter 3**), whether combining implicit measures in a model improves the sensitivity (**Chapter 4**), and how emotional context affects pleasantness of novel and familiar drinks as measured using both explicit and implicit measures (**Chapter 5**).

The literature review in **Chapter 2** shows that previous studies predominantly rely on explicit ratings and questionnaires or a combination of redundant explicit measures to assess food-evoked emotions. We recommend researchers to use a combination of physiological, behavioral, and cognitive measures that covers all three emotional processing levels by providing a 3 x 3 complementary toolbox table.

The study in **Chapter 3** provides an overview of the sensitivity of simultaneously measured explicit and implicit responses to regular drinks and a vinegar solution. It shows that besides explicit measures of valence (pleasantness) and arousal, most implicit measures are sensitive to large differences in affective experience (regular drinks and the vinegar solution), where sip size had the largest discriminative power. HR was sensitive even to subtle differences in affective experience as indicated by a correlation between explicit arousal ratings and HR. **Chapter 4** shows that a predicting model trained by using implicit responses (especially sip size and HR) to the vinegar solution as a ground truth high arousal-low valence drink, and to regular drinks as low arousal-high valence drinks, result in quite accurate estimates of explicit arousal scores assigned to subtle different regular drinks. In contrast, the traditional approach of using explicit valence and arousal ratings as a ground truth to train the model failed to provide sophisticated predictions.

The study described in **Chapter 5** examines the effect of emotional context on the pleasantness of novel and familiar foods using different types of measures. It shows that familiar soups were rated more positively than novel soups in a negative emotional condition whereas both soups were equally rated in a positive emotional state. The effect of emotional state was robust over at least one week. Besides explicit ratings, implicit behavioral measures (sip size and willingness-to-take-home) indicate similar patterns of effects of soup, state, and session.

The second part of the thesis addresses measures of food-evoked emotion in a cross-cultural context. It describes the development and validation of the food image database CROss-CUltural Food Image Database (CROCUFID)" (**Chapter 6**), the development and validation of a language independent tool with Emoji to evaluate food-evoked emotions (so called EmojiGrid) (**Chapter 7**), to what extent this EmojiGrid can reliably assess food experience across cultures (**Chapter 8**), and the potential of explicit and implicit measures to cross-culturally assess food-evoked emotions (**Chapter 9**).

Chapter 6 introduces a new food image database (CROCUFID) to facilitate cross-cultural study in food research to meet the demand for non-western and low valence food images. Validation data for the CROCUFID food images were obtained from a total of 805 participants (ranging in age between 18 and 70) with different cultural background (United Kingdom, United states, and Japan). CROCUFID now contains more than 980 standardized food and non-food images, covering some of Asian and Western cuisines. The database is still growing through contributions of researchers around the world.

In **Chapter 7 and 8**, a new graphical tool with Emojis called "EmojiGrid" is introduced and validated through cross-cultural studies. The study in **Chapter 7** shows that each of 16 self-designed Emojis conveys the intended degrees of valence and arousal using a correlation analysis between these Emojis and a conventional graphical tool called SAM scale. Also, this chapter shows that the EmojiGrid is an intuitive and efficient graphical tool to capture the affective dimensions of valence and arousal for food image stimuli. **Chapter 8** further explores the potential of EmojiGrid for cross-cultural studies in respondents with different mother tongues (respondents from Germany, the Netherlands, United Kingdom and Japan). This was done through an online experiment including 60 food images covering a wide range of valence. The study shows that as expected, the EmojiGrid produces the nomothetic U-shaped relation between valence and arousal ratings on the same food image set for all four different countries, where the pattern of the Asian country (Japan) differs from the other, western countries. The EmojiGrid appears to be a valid and language-independent self-report tool for cross-cultural research on food-related emotions.

Chapter 9 investigates the potential of implicit measures to capture food-evoked emotions in across cultures. Participants from Thailand (representative for Middle Response Style (MRS) in explicit measures) and the Netherlands (representative for an Extreme Response Style (ERS) in explicit measures) were presented with food images and drinks from their own and other culture. In addition, images of universal and low valence (disgusting) food were shown. This study shows the expected explicit response bias in Thai and Dutch participants, while implicit measures of HR (in case of food images) and sip size (in case of drinks) followed the prior expectations of genuine food experience. The study suggests that physiological responses may complement self-reports in that they are insensitive to cultural response bias that could lead to misunderstanding across cultures.

The thesis ends with **Chapter 10**, discussing the main findings of the presented studies as a whole, as well as the important aspects from a commercial point of view. We emphasize that besides explicit measures, it is recommended to combine implicit (neuro)physiological and behavioral measures to better understand food-evoked emotions. Also, we newly developed a validated food image database (CROCUFID) and an intuitive language-independent affective evaluation tool (EmojiGrid) to facilitate cross-cultural studies in food research. Throughout the thesis, implicit measures and combining implicit measures show their potential to capture food-evoked emotions, especially arousal, without asking explicitly. Among them, HR and sip size may be the most promising implicit responses to measure emotions evoked by food experiences. However, new devices that are able to remotely measure implicit (neuro)physiological and behavioral responses without interrupting individuals' daily life are eagerly anticipated for further investigations on food-evoked emotions.

Acknowledgements

First and foremost I am extremely grateful to my promotor, Prof. Jan van Erp and my supervisor, Dr. Anne-Marie Brouwer for their invaluable advice, continuous support for long time, and patience during my PhD study. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research. Jan was an ally throughout the project and always gave me kind words. Anne-Marie was always surrounding me at the office and brought me to get acquainted with various international internship students and your colleagues. Also, I'd like to thank Anne-Marie for your supports of our life in the Netherlands including my family. There are many, many sweet memories with you and your family during my stay in the Netherlands, especially the scenes in several different restaurants and bars we drink together and the scene in "game evening" at your home. Also, I had nice memories to participate together in various conferences. These unforgettable scenes made my study and life in the Netherlands a wonderful time. I'm for sure that I would have not make it to the end without your supports and encouragement.

I would also like to thank Dr. Alexander Toet for your support writing draft manuscripts and for your-never-give-up attitude on writing manuscript, which strongly encouraged me to write several research manuscripts. Several EmojiGrid validation studies and cross-cultural studies would not go so smoothly without you. It was fun and a good memory for me to have scotch whiskey with you after a workshop held in Glasgow and a conference in Edinburg. I also never forget your smile that you and your wife came to my home to hold my one month daughter born in the Netherlands. I appreciate Dr. Victor Kallen for always encouraging me with your positive mind and words. The new validated EmojiGrid was never developed without your fruitful idea. Dr. Annelies Dijk-Stroeve, thank you for your supports to manage the project as a project manager. Furthermore, I am also pleased to say thank you to Dr. Maarten Hogervorst, Mr. Martin van Schaik, Mr. Ivo Stuldreher, Mr. Tim van den Broek, and Mr. Jan Ubbo Baardewijk for their technical supports and contributions on data analysis and discussions of the analyzed data.

During my PhD period, many internship students kindly supported me for conducting the several experiments, Wouter, Esther, Benjamin, Inge, Sophia, and Anne. Wouter, Esther, and Benjamin, you were my first internship students for the project, and I remember we had fun preparing several different drink samples together. Inge, thank you for preparing food samples and collecting food images by taking pictures of them together. It was fun to prepare fresh and molded foods together. Anne, I appreciate your effort to prepare the experiment and to collect another Dutch food images together. Thank you all again for your supports together with me. I remember we chat a lot about the different culture between western countries and Japan during lunch and dinner together. It was all fun for me and nice memories.

To Mr. Wim van Hartingsveldt and Ms. Kaori Nishide, I appreciate both of you to help me when I just arrived at the Netherlands first time. Wim, I remember I met you first time in front of Amersfoort central station as soon as I came to the Netherlands. Also, I appreciate you to keep watching me in terms of the project progress and my life in the Netherlands. Furthermore, I would

like to thank all the business managers in relevant departments in TNO especially for understanding my situation and allowing me to hold my position as a guest researcher in TNO.

I kindly appreciate Prof. Suwimon Keeratipibul, Dr. Punnida Techaruvichit, and Dr. Panita Ngamchuachit for your supports lending us the room in your laboratory in Chulalongkorn University and recruiting Thai participants for our cross-cultural study there. Also, I thank Dr. Nattapong Thammasan for supporting the cross-cultural study together in Thailand and for collecting Thai food images in a Thai restaurant in the Netherlands. It was an amazing experience in Bangkok in Thailand with you, and Thai foods there were amazing to me.

I am also pleased to appreciate my boss, Dr. Asahi Matsuyama for your patience during my stay in the Netherlands. Your suggestions and advices on the project progress during my work in the Netherlands were fruitful, and I haven't finished publishing research outcomes from the project as many research articles without your financial supports and fruitful suggestions.

I am grateful to my parents, siblings, friends and acquaintances who remembered me in their prayers for the ultimate success. I especially thank my parents to mentally support me through online chatting with your positive words. I consider myself nothing without them. They gave me enough moral support, encouragement and motivation to accomplish the personal goals.

Finally, I would like to express my gratitude to my beloved family staying together in the Netherlands, my wife, Aya Kaneko, and my children, Daigo Kaneko and Rin Kaneko. Aya, you always support me for my daily life a lot and keep pushing me to continue working in the Netherlands, and the carefree smile of my children at home for me encourage a lot to motivate working every time I was a bit depressed during the really difficult situation due to the corona pandemic all over the world. Without their tremendous understanding and encouragement and children's in the past few years, it would be impossible for me to complete my study.

References

Adams, J., Williams, A., Lancaster, B., and Foley, M. (2007). "Advantages and uses of Check-All-That-Apply response compared to traditional scaling of attributes for salty snacks", in: *7th Pangborn sensory science symposium.* (Minneapolis, USA).

Agarwal, S., and Xavier, M.J. (2015). "Innovations in consumer science: applications of neuro-scientific research tools," in *Adoption of innovation: Balancing internal and external stakeholders in the marketing of innovation,* eds. A. Brem & É. Viardot. (Cham: Springer International Publishing), 25-42.

Aldridge, V., Dovey, T.M., and Halford, J.C. (2009). The role of familiarity in dietary development. *Developmental Review* 29**,** 32-44.

Alismail, S., and Zhang, H. (2018). "The use of emoji in electronic user experience questionnaire: An exploratory case study", in: *51st Hawaii International Conference on System Sciences.* (Waikoloa Village, HI: Urban Center for Computation and Data (URBANCCD)).

Aluja, A., Balada, F., Blanco, E., Lucas, I., and Blanch, A. (2018). Startle reflex modulation by affective face “Emoji” pictographs. *Psychological Research* Online first.

Anderson, A.K., Christoff, K., Stappen, I., Panitz, D., Gharemani, D.G., Glover, G., Gabrieli, J.D.E., and Soberl, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience* 6**,** 196-202.

Anderson, A.K., Wais, P.E., and Gabrieli, J.D. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences* 103**,** 1599-1604.

Anderson, E.C., Wormwood, J., Barrett, L.F., and Quigley, K.S. (2019). Vegetarians’ and omnivores’ affective and physiological responses to images of food. *Food quality and preference* 71**,** 96-105.

Andrienko, Y.A., Brilliantov, N.V., and Kurths, J. (2000). Complexity of two-dimensional patterns. *The European Physical Journal B - Condensed Matter and Complex Systems* 15**,** 539-546.

Anolli, L., Mantovani, F., Confalonieri, L., Ascolese, A., and Peveri, L. (2010). Emotions in serious games: From experience to assessment. *International Journal of Emerging Technologies in Learning* 5**,** 7-16.

Anttonen, J., and Surakka, V. (Year). "Emotions and heart rate while sitting on a chair", in: *Proceedings of the SIGCHI conference on Human factors in computing systems*), 491-499.

Appelhans, B.M., and Luecken, L.J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of general psychology* 10**,** 229.

Ares, G. (2018). Methodological issues in cross-cultural sensory and consumer research. *Food Quality and Preference* 64**,** 253-263.

Ares, G., Barreiro, C., Deliza, R., Giménez, A., and Gambaro, A. (2010). Application of a Check-All-That-Apply question to the development of chocolate milk desserts. *Journal of Sensory Studies***,** 67-86.

Ares, G., Bruzzone, F., Vidal, L., Cadena, R.S., Giménez, A., Pineau, B., Hunter, D.C., Paisley, A.G., and Jaeger, S.R. (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.

Ares, G., Giménez, A., and Gambaro, A. (2008). Understanding consumers' perception of conventional and functional yogurts using word association and hard laddering. *Food Quality and Preference* 19**,** 636-643.

Ares, G., and Jaeger, S.R. (2017). A comparison of five methodological variants of emoji questionnaires for measuring product elicited emotional associations: An application with seafood among Chinese consumers. *Food Research International* 99**,** 216-228.

Aronoff, J., Barclay, A.M., and Stevenson, L.A. (1988). The recognition of threatening facial stimuli. *Journal of Personality and Social Psychology* 54**,** 647-655.

Aust, F., Diedenhofen, B., Ullrich, S., and Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods* 45**,** 527-535.

Awazu, S. (2013). Affective information influences affective evaluation but not affective expectation. *PSYCHOLOGIA* 56**,** 237-245.

Bach, A.J., Stewart, I.B., Minett, G.M., and Costello, J.T. (2015). Does the technique employed for skin temperature assessment alter outcomes? A systematic review. *Physiological measurement* 36**,** R27.

Bagozzi, R.P., Gopinath, M., and Nyer, P.U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science* 27**,** 184-206.

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

Barla, A., Franceschi, E., Odone, F., and Verri, A. (2002). "Image Kernels," in *Pattern Recognition with Support Vector Machines: First International Workshop, SVM 2002 Niagara Falls, Canada, August 10, 2002 Proceedings,* eds. S.-W. Lee & A. Verri. (Berlin, Heidelberg: Springer Berlin Heidelberg), 83-96.

Becker, D., Jostmann, N.B., Wiers, R.W., and Holland, R.W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite* 85**,** 58-65.

Bekkedal, M.Y., Rossi Iii, J., and Panksepp, J. (2011). Human brain EEG indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neuroscience & Biobehavioral Reviews* 35**,** 1959-1970.

Bell, A.J., and Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation* 7**,** 1129-1159.

Bell, L., Vogt, J., Willemse, C., Routledge, T., Butler, L.T., and Sakaki, M. (2018). Beyond self-report: A review of physiological and neuroscientific methods to investigate consumer behavior. *Frontiers in psychology* 9**,** 1655.

Bellisle, F. (2009). How and why should we study ingestive behaviors in humans? *Food quality and preference* 20**,** 539-544.

Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods* 190**,** 80-91.

Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., and Holley, A. (2002). Autonomic nervous system responses to odours: the role of pleasantness and arousal. *Chemical Senses* 27**,** 703-709.

Bercea, M.D. (2013). "Anatomy of methodologies for measuring consumer behavior in neuromarketing research", in: *Lupcon Center for Business Research (LCBR) European Marketing Conference 2012.* Lupcon Center for Business Research).

Berlyne, D.E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology* 55**,** 289-296.

Berntson, G.G., Thomas Bigger, J., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., and Stone, P.H. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34**,** 623-648.

Berridge, K.C. (1996). Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews* 20**,** 1-25.

Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior* 81**,** 179-209.

Berridge, K.C., Ho, C.-Y., Richard, J.M., and Difeliceantonio, A.G. (2010). The tempted brain eats: pleasure and desire circuits in obesity and eating disorders. *Brain research* 1350**,** 43-64.

Berthoud, H.-R., and Zheng, H. (2012). Modulation of taste responsiveness and food preference by obesity and weight loss. *Physiology & Behavior* 107**,** 527-532.

Betella, A., and Verschure, P.F.M.J. (2016). The Affective Slider: A digital self-assessment scale for the measurement of human emotions. *PLOS ONE* 11**,** e0148037.

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

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

Birch, L.L., Zimmerman, S.I., and Hind, H. (1980). The influence of social-affective context on the formation of children's food preferences. *Child development***,** 856-861.

Bitner, M.J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *The Journal of Marketing***,** 57-71.

Blechert, J., Goltsche, J.E., Herbert, B.M., and Wilhelm, F.H. (2014a). Eat your troubles away: Electrocortical and experiential correlates of food image processing are related to emotional eating style and emotional state. *Biological Psychology* 96**,** 94-101.

Blechert, J., Meule, A., Busch, N.A., and Ohla, K. (2014b). Food-pics: an image database for experimental research on eating and appetite. *Frontiers in Psychology* 5**,** 617.

Blechert, J., Meule, A., Busch, N.A., and Ohla, K. (2014c). Food-pics: an image database for experimental research on eating and appetite. *Frontiers in Psychology* 5**,** Article 617.

Bond, A., and Lader, M. (1974). The use of analogue scales in rating subjective feelings. *British Journal of Medical Psychology* 47**,** 211-218.

Bosch, A., Zisserman, A., and Munoz, X. (2007). "Representing shape with a spatial pyramid kernel", in: *6th ACM international conference on Image and video retrieval.* (Amsterdam, The Netherlands: ACM).

Bossard, L., Guillaumin, M., and Van Gool, L. (Year). "Food-101 – Mining discriminative components with random forests": Springer International Publishing), 446-461.

Boster, J.S. (2005a). "Emotion categories across languages," in *Handbook of categorization in cognitive science*. Elsevier), 313-352.

Boster, J.S. (2005b). "Emotion Categories Across Languages A2 - Cohen, Henri," in *Handbook of Categorization in Cognitive Science (Second Edition),* ed. C. Lefebvre. (San Diego, CA: Elsevier), 313-352.

Boyce, J.M., and Shone, G.R. (2006). Effects of ageing on smell and taste. *Postgraduate Medical Journal* 82**,** 239-241.

Bradley, M.M., Cuthbert, B.N., and Lang, P.J. (1990). Startle reflex modification: Emotion or attention? *Psychophysiology* 27**,** 513-522.

Bradley, M.M., Greenwald, M.K., Petry, M.C., and Lang, P.J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of experimental psychology: Learning, Memory, and Cognition* 18**,** 379-390.

Bradley, M.M., and Lang, P.J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25**,** 49-59.

Bradley, M.M., and Lang, P.J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37**,** 204-215.

Braun, J., Amirshahi, S.A., Denzler, J., and Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology* 4**,** 808.

Braun, M.H., Pradana, G.A., Buchanan, G., Cheok, A.D., Velasco, C., Spence, C., Aduriz, A.L., Gross, J., and Lasa, D. (2016). Emotional priming of digital images through mobile telesmell and virtual food. *International Journal of Food Design* 1**,** 29-45.

Brockmeyer, T., Hahn, C., Reetz, C., Schmidt, U., and Friederich, H.-C. (2015). Approach bias and cue reactivity towards food in people with high versus low levels of food craving. *Appetite* 95**,** 197-202.

Broekens, J., and Brinkman, W.P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies* 71**,** 641-667.

Brouwer, A.-M., Hogervorst, M., Grootjen, M., Van Erp, J., and Zandstra, E. (2017a). Neurophysiological responses during cooking food associated with different emotions. *Food Quality and Preference*.

Brouwer, A.-M., and Hogervorst, M.A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in neuroscience* 8**,** 224.

Brouwer, A.-M., Hogervorst, M.A., Van Erp, J.B., Grootjen, M., Van Dam, E., and Zandstra, E.H. (2019). Measuring cooking experience implicitly and explicitly: Physiology, facial expression and subjective ratings. *Food quality and preference* 78**,** 103726.

Brouwer, A.-M., Van Beers, J.J., Sabu, P., Stuldreher, I.V., Zech, H.G., and Kaneko, D. (2021). Measuring Implicit Approach–Avoidance Tendencies towards Food Using a Mobile Phone Outside the Lab. *Foods* 10**,** 1440.

Brouwer, A.-M., Van Beurden, M., Nijboer, L., Derikx, L., Binsch, O., Gjaltema, C., and Noordzij, M. (Year). "A comparison of different electrodermal variables in response to an acute social stressor", in: *International Workshop on Symbiotic Interaction*: Springer), 7-17.

Brouwer, A.-M., Van Den Broek, T., Hogervorst, M., Kaneko, D., Toet, A., Kallen, V., and Van Erp, J. (2020). Estimating affective taste experience using combined implicit behavioral and neurophysiological measures. *IEEE Transactions on Affective Computing*.

Brouwer, A.-M., Zander, T.O., Van Erp, J.B., Korteling, J.E., and Bronkhorst, A.W. (2015a). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience* 9.

Brouwer, A., Hogervorst, M., Grootjen, M., Van Erp, J., and Zandstra, E. (2017c). Neurophysiological responses during cooking food associated with different emotions. *Food Quality and Preference*.

Brouwer, A., Hogervorst, M., Grootjen, M., Van Erp, J., and Zandstra, E. (2017d). Neurophysiological responses during cooking food associated with different emotions. *Food Quality and Preference***,** 307-316.

Brouwer, A., Hogervorst, M., Reuderink, B., Van Der Werf, Y., and Van Erp, J. (2015b). Physiological signals distinguish between reading emotional and non-emotional sections in a novel. *Brain-Computer Interfaces* 2**,** 76-89.

Brouwer, A., Zander, T., Van Erp, J., Korteling, J., and Bronkhorst, A. (2015c). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience* 9.

Brunyé, T.T., Hayes, J.F., Mahoney, C.R., Gardony, A.L., Taylor, H.A., and Kanarek, R.B. (2013). Get in My Belly: Food Preferences Trigger Approach and Avoidant Postural Asymmetries. *PLoS ONE* 8**,** e72432.

Burger, K.S., Cornier, M.A., Ingebrigtsen, J., and Johnson, S.L. (2011). Assessing food appeal and desire to eat: the effects of portion size & energy density. *The International Journal of Behavioral Nutrition and Physical Activity* 8**,** 101-101.

Cabanac, M. (2002). What is emotion? *Behavioural Processes* 60**,** 69-83.

Cacioppo, J.T., and Gardner, W.L. (1999). Emotion. *Annual review of psychology* 50**,** 191-214.

Cacioppo, J.T., Priester, J.R., and Berntson, G.G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of personality and social psychology* 65**,** 5.

Cai, H., and Lin, Y. (2011). Modeling of operators' emotion and task performance in a virtual driving environment. *International Journal of Human-Computer Studies* 69**,** 571-586.

Camm, A.J., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H., and Kleiger, R. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* 93**,** 1043-1065.

Cano, M.E., Class, Q.A., and Polich, J. (2009). Affective valence, stimulus attributes, and P300: Color vs. black/white and normal vs. scrambled images. *International Journal of Psychophysiology* 71**,** 17-24.

Caporale, G., Policastro, S., Tuorila, H., and Monteleone, E. (2009). Hedonic ratings and consumption of school lunch among preschool children. *Food Quality and Preference* 20**,** 482-489.

Cardello, A.V., Meiselman, H.L., Schutz, H.G., Craig, C., Given, Z., Lesher, L.L., and Eicher, S. (2012). Measuring emotional responses to foods and food names using questionnaires. *Food Quality and Preference* 24**,** 243-250.

Cepeda-Benito, A., Fernandez, M.C., and Moreno, S. (2003). Relationship of gender and eating disorder symptoms to reported cravings for food: construct validation of state and trait craving questionnaires in Spanish. *Appetite* 40**,** 47-54.

Charbonnier, L., Van Der Laan, L.N., Viergever, M.A., and Smeets, P.a.M. (2015). Functional MRI of challenging food choices: Forced choice between equally liked high- and low-calorie foods in the absence of hunger. *PLOS ONE* 10**,** e0131727.

Charbonnier, L., Van Meer, F., Van Der Laan, L.N., Viergever, M.A., and Smeets, P.a.M. (2016). Standardized food images: A photographing protocol and image database. *Appetite* 96**,** 166-173.

Chen, C., Lee, S.-Y., and Stevenson, H.W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science* 6**,** 170-175.

Chen, J., and Ngo, C.-W. (2016). "Deep-based ingredient recognition for cooking recipe retrieval", in: *2016 ACM on Multimedia Conference.* (Amsterdam, The Netherlands: ACM).

Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R., and Yang, J. (2009). "PFID: Pittsburgh fast-food image dataset", in: *6th IEEE International Conference on Image Processing (ICIP).*).

Chen, X., Zhou, H., and Diao, L. (2017). ChineseFoodNet: A large-scale Image dataset for Chinese food recognition. *arXiv preprint arXiv:1705.02743*.

Chen, Y.-C., Woods, A.T., and Spence, C. (2018a). Sensation transference from plateware to food: The sounds and tastes of plates. *International Journal of Food Design* 3**,** 41-62.

Chen, Y., Gao, Q., Lv, Q., Qian, N., and Ma, L. (2018b). Comparing measurements for emotion evoked by oral care products. *International Journal of Industrial Ergonomics* 66**,** 119-129.

Cho, S., Han, A., Taylor, M.H., Huck, A.C., Mishler, A.M., Mattal, K.L., Barker, C.A., and Seo, H.-S. (2015). Blue lighting decreases the amount of food consumed in men, but not in women. *Appetite* 85**,** 111-117.

Choi, Y.K., and Taylor, C.R. (2014). How do 3-dimensional images promote products on the Internet? *Journal of Business Research* 67**,** 2164-2170.

Christie, I.C., and Friedman, B.H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach. *International Journal of Psychophysiology* 51**,** 143-153.

Chung, L., Chung, S.-J., Kim, J.-Y., Kim, K.-O., O'mahony, M., Vickers, Z., Cha, S.-M., Ishii, R., Baures, K., and Kim, H.-R. (2012). Comparing the liking for Korean style salad dressings and beverages between US and Korean consumers: Effects of sensory and non-sensory factors. *Food Quality and Preference* 26**,** 105-118.

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

Clark, J.E. (1998). Taste and flavour: their importance in food choice and acceptance. *Proceedings of the Nutrition Society* 57**,** 639-643.

Clarke, T., and Costall, A. (2008). The emotional connotations of color: A qualitative investigation. *Color Research & Application* 33**,** 406-410.

Codispoti, M., and De Cesarei, A. (2007). Arousal and attention: Picture size and emotional reactions. *Psychophysiology* 44**,** 680-686.

Comber, R., Choi, J.H.-J., Hoonhout, J., and O'hara, K. (2014a). Designing for human–food interaction: An introduction to the special issue on ‘food and interaction design’. *International Journal of Human-Computer Studies* 72**,** 181-184.

Comber, R., Choi, J.H.-J., Hoonhout, J., and O'hara, K. (2014b). Editorial: Designing for human-food interaction: An introduction to the special issue on 'food and interaction design'. *International Journal of Human-Computer Studies* 72**,** 181-184.

Comber, R., Hoonhout, J., Halteren, A.V., Moynihan, P., and Olivier, P. (2013). "Food practices as situated action: exploring and designing for everyday food practices with households", in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* (Paris, France: ACM).

Coppin, G., and Sander, D. (2016). "Theoretical approaches to emotion and its measurement," in *Emotion measurement*. Elsevier), 3-30.

Corchs, S.E., Ciocca, G., Bricolo, E., and Gasparini, F. (2016). Predicting complexity perception of real world images. *PLOS ONE* 11**,** e0157986.

Cordonnier, S., and Delwiche, J. (2008). An alternative method for assessing liking: Positional relative rating versus the 9-point hedonic scale. *Journal of Sensory Studies* 23**,** 284-292.

Cowie, R., Douglas-Cowie, E., Savvidou, S., Mcmahon, E., M., S., and Schröder, M. (2000). "FEELTRACE: An instrument for recording perceived emotion in real time", in: *ISCA Workshop on Speech and Emotion: Developing a Conceptual Framework.* (Newcastle, N. Ireland: Textflow).

Crisinel, A.-S., Cosser, S., King, S., Jones, R., Petrie, J., and Spence, C. (2012). A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background. *Food Quality and Preference* 24**,** 201-204.

Cunningham, W.A., Raye, C.L., and Johnson, M.K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience* 16**,** 1717-1729.

Curia, A.V., Hough, G., MartíNez, M.C., and Margalef, M.I. (2001). How Argentine consumers understand the Spanish translation of the 9-point hedonic scale. *Food Quality and Preference* 12**,** 217-221.

Daft, R.L., and Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management science* 32**,** 554-571.

Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection", in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005).* IEEE).

Dalenberg, J.R., Gutjar, S., Ter Horst, G.J., De Graaf, K., Renken, R.J., and Jager, G. (2014). Evoked emotions predict food choice. *PLoS ONE* 9**,** e115388.

Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the internet.* London, New York: Bloomsbury Publishing.

Danner, L., Haindl, S., Joechl, M., and Duerrschmid, K. (2014a). Facial expressions and autonomous nervous system responses elicited by tasting different juices. *Food Research International* 64**,** 81-90.

Danner, L., Sidorkina, L., Joechl, M., and Duerrschmid, K. (2014b). 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.

Davies, S.R., El-Deredy, W., Zandstra, E.H., and Blanchette, I. (2012). Evidence for the role of cognitive resources in flavour–flavour evaluative conditioning. *The Quarterly Journal of Experimental Psychology* 65**,** 2297-2308.

Dawson, M.E., Schell, A.M., and Filion, D.L. (2017). The electrodermal system.

De Araujo, I.E., Rolls, E.T., Velazco, M.I., Margot, C., and Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron* 46**,** 671-679.

De Wijk, R., and Boesveldt, S. (2016). "Responses of the Autonomic Nervous System to Flavors," ed. B. Piqueras-Fiszman.).

De Wijk, R., Kaneko, D., Dijksterhuis, G., Van Zoggel, M., Schiona, I., Visalli, M., and Zandstra, E. (2019). Food perception and emotion measured over time in-lab and in-home. *Food Quality and Preference* 75**,** 170-178.

De Wijk, R.A., He, W., Mensink, M.G.J., Verhoeven, R.H.G., and De Graaf, C. (2014). ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks. *PLoS ONE* 9**,** 1-9.

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

De Wijk, R.A., and Noldus, L.P. (2021). Using implicit rather than explicit measures of emotions. *Food Quality and Preference* 92**,** 104125.

De Wijk, R.A., Ushiama, S., Ummels, M.J., Zimmerman, P.H., Kaneko, D., and Vingerhoeds, M.H. (2021). Effect of Branding and Familiarity of Soy Sauces on Valence and Arousal as Determined by Facial Expressions, Physiological Measures, Emojis, and Ratings. *Frontiers in Neuroergonomics* 2**,** 13.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134**,** 9-21.

Den Uijl, L.C., Jager, G., De Graaf, C., Meiselman, H.L., and Kremer, S. (2016a). Emotion, olfaction, and age: A comparison of self-reported food-evoked emotion profiles of younger adults, older normosmic adults, and older hyposmic adults. *Food Quality and Preference* 48**,** 199-209.

Den Uijl, L.C., Jager, G., Zandstra, E.H., De Graaf, C., and Kremer, S. (2016b). Self-reported food-evoked emotions of younger adults, older normosmic adults, and older hyposmic adults as measured using the PrEmo2 tool and the Affect Grid. *Food Quality and Preference* 51**,** 109-117.

Desmet, P.M.A., Hekkert, P., and Jacobs, J.J. (2000a). "When a car makes you smile: Development and application of an instrument to measure product emotions", in: *Advances in Consumer Research.* (eds.) S.J. Hoch & R.J. Meyer. Association for Consumer Research).

Desmet, P.M.A., Hekkert, P., and Jacobs, J.J. (Year). "When a car makes you smile: Development and application of an instrument to measure product emotions", eds. S.J. Hoch & R.J. Meyer), 111-117.

Desmet, P.M.A., Hekkert, P., and Jacobs, J.J. (Year). "When a car makes you smile: Development and application of an instrument to measure product emotions", in: *Advances in Consumer Research*, eds. S.J. Hoch & R.J. Meyer (Provo, UT, USA: Association for Consumer Research), 111-117.

Desmet, P.M.A., and Schifferstein, H.N.J. (2008). Sources of positive and negative emotions in food experience. *Appetite* 50**,** 290-301.

Desmet, P.M.A., Vastenburg, M.H., and Romero, N. (2016). Mood measurement with Pick-A-Mood: review of current methods and design of a pictorial self-report scale. *Journal of Design Research* 14**,** 241-279.

Desmet, P.M.A., Vastenburg, M.H., Van Bel, D.T., and Romero Herrera, N.A. (Year). "Pick-A-Mood; development and application of a pictorial mood-reporting instrument", in: *Out of Control: Proceedings of the 8th International Conference on Design and Emotion, London, UK, 11-14 September 2012*, eds. J. Brassett, P. Hekkert, G. Ludden, M. Malpass & J. Mcdonnell).

Di Pellegrino, G., Magarelli, S., and Mengarelli, F. (2011). Food pleasantness affects visual selective attention. *Q J Exp Psychol (Hove)* 64**,** 560-571.

Donderi, D. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception* 35**,** 823-835.

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

Dos Reis, J.C., Bonacin, R., Hornung, H.H., and Baranauskas, M.C.C. (2018). Intenticons: Participatory selection of emoticons for communication of intentions. *Computers in Human Behavior* 85**,** 146-162.

Einöther, S.J.L., Baas, M., Rowson, M., and Giesbrecht, T. (2015). Investigating the effects of tea, water and a positive affect induction on mood and creativity. *Food Quality and Preference* 39**,** 56-61.

Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin* 115**,** 268-287.

Ekman, P. (2003). *Emotions revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life.* New York, NY, USA: Henry Holt and Comnapny, LLC.

Ekman, P., and Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17**,** 124-129.

Ekman, P., and Friesen, W.V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues.* Cambridge, MA: Malor Books.

Ekman, P., Friesen, W.V., and Tomkins, S.S. (1971). Facial affect scoring technique: A first validity study. *Semiotica* 3**,** 37-58.

Ekman, P., and Rosenberg, E.L. (2004). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), 2nd ed.* New York, USA: Oxford University Press.

Elder, A.M. (2018). What words can't say: Emoji and other non-verbal elements of technologically-mediated communication. *Journal of Information, Communication and Ethics in Society* 16**,** 2-15.

Elliot, A.J., and Covington, M.V. (2001). Approach and avoidance motivation. *Educational Psychology Review* 13**,** 73-92.

Esteves, F., Arriaga, P., Carneiro, P., and Flykt, A. (2010). Emotional responses (verbal and psychophysiological) to pictures of food stimuli. *Psicologia* 24**,** 89-111.

Farinella, G.M., Allegra, D., and Stanco, F. (Year). "A benchmark dataset to study the representation of food images": Springer International Publishing), 584-599.

Fenko, A., Backhaus, B.W., and Van Hoof, J.J. (2015a). The influence of product- and person-related factors on consumer hedonic responses to soy products. *Food Quality and Preference* 41**,** 30-40.

Fenko, A., Backhaus, B.W., and Van Hoof, J.J. (2015b). The influence of product-and person-related factors on consumer hedonic responses to soy products. *Food quality and preference* 41**,** 30-40.

Ferber, C., and Cabanac, M. (1987). Influence of noise on gustatory affective ratings and preference for sweet or salt. *Appetite* 8**,** 229-235.

Fiegel, A., Meullenet, J.F., Harrington, R.J., Humble, R., and Seo, H.S. (2014). Background music genre can modulate flavor pleasantness and overall impression of food stimuli. *Appetite* 76**,** 144-152.

Fleureau, J., Guillotel, P., and Quan, H. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing* 3**,** 379-385.

Flint, A., Raben, A., Blundell, J.E., and Astrup, A. (2000). Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *International journal of obesity and related metabolic disorders* 24**,** 38-48.

Foroni, F., Pergola, G., Argiris, G., and Rumiati, R. (2013). The FoodCast research image database (FRIDa). *Frontiers in Human Neuroscience* 7**,** 1-19.

Foroni, F., Pergola, G., and Rumiati, R.I. (2016). Food color is in the eye of the beholder: the role of human trichromatic vision in food evaluation. *Scientific Reports* 6.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C.J., and Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art. *British Journal of Psychology* 102**,** 49-70.

Frühholz, S., Jellinghaus, A., and Herrmann, M. (2011). Time course of implicit processing and explicit processing of emotional faces and emotional words. *Biological Psychology* 87**,** 265-274.

Gable, P., and Harmon-Jones, E. (2008). Relative left frontal activation to appetitive stimuli: Considering the role of individual differences. *Psychophysiology* 45**,** 275-278.

Gable, P.A., and Harmon-Jones, E. (2009). Postauricular reflex responses to pictures varying in valence and arousal. *Psychophysiology* 46**,** 487-490.

Gallo, K.E., Swaney-Stueve, M., and Chambers, D.H. (2017). A focus group approach to understanding food-related emotions with children using words and emojis. *Journal of Sensory Studies* 32**,** e12264.

Garcia-Burgos, D., and Zamora, M.C. (2013). Facial affective reactions to bitter-tasting foods and body mass index in adults. *Appetite* 71**,** 178-186.

Geier, U., Büssing, A., Kruse, P., Greiner, R., and Buchecker, K. (2016). Development and Application of a Test for Food-Induced Emotions. *PloS one* 11**,** e0165991.

Genschow, O., Reutner, L., and Wänke, M. (2012). The color red reduces snack food and soft drink intake. *Appetite* 58**,** 699-702.

Gibson, E.L. (2006). Emotional influences on food choice: sensory, physiological and psychological pathways. *Physiology & Behavior* 89**,** 53-61.

Gil, S., Rousset, S., and Droit-Volet, S. (2009). How liked and disliked foods affect time perception. *Emotion* 9**,** 457-463.

Giomo, C.J. (1993). An experimental study of children's sensitivity to mood in music. *Psychology of Music* 21**,** 141-162.

Goedhart, A.D., Van Der Sluis, S., Houtveen, J.H., Willemsen, G., and De Geus, E.J. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology* 44**,** 203-215.

Gomes, L.M., Silva, R.G., Melo, M., Silva, N.N., Vanderlei, F.M., Garner, D.M., De Abreu, L.C., and Valenti, V.E. (2016). Effects of Effortful Swallow on Cardiac Autonomic Regulation. *Dysphagia* 31**,** 188-194.

Gorini, A., Griez, E., Petrova, A., and Riva, G. (2010a). Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry* 9**,** 30.

Gorini, A., Griez, E., Petrova, A., and Riva, G. (2010b). Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry* 9**,** 1-10.

Gosling, S.D., Vazire, S., Srivastava, S., and John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American psychologist* 59**,** 93-104.

Grabenhorst, F., and Rolls, E.T. (2008). Selective attention to affective value alters how the brain processes taste stimuli. *European Journal of Neuroscience* 27**,** 723-729.

Grabenhorst, F., Rolls, E.T., and Bilderbeck, A. (2008). How cognition modulates affective responses to taste and flavor: Top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cerebral Cortex* 18**,** 1549-1559.

Greenwald, A.G., Klinger, M.R., and Liu, T.J. (1989). Unconscious processing of dichoptically masked words. *Memory & cognition* 17**,** 35-47.

Greenwald, A.G., Mcghee, D.E., and Schwartz, J.L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology* 74**,** 1464.

Griffioen-Roose, S., Smeets, P.A., Weijzen, P.L., Van Rijn, I., Van Den Bosch, I., and De Graaf, C. (2013a). Effect of replacing sugar with non-caloric sweeteners in beverages on the reward value after repeated exposure. *PLoS One* 8**,** e81924.

Griffioen-Roose, S., Smeets, P.a.M., Weijzen, P.L.G., Van Rijn, I., Van Den Bosch, I., and De Graaf, C. (2013b). Effect of replacing sugar with non-caloric sweeteners in beverages on the reward value after repeated exposure. *PLOS ONE* 8**,** e81924.

Grossman, P., and Taylor, E.W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological psychology* 74**,** 263-285.

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

Gutjar, S., De Graaf, C., Kooijman, V., De Wijk, R.A., Nys, A., Ter Horst, G.J., and Jager, G. (2015b). The role of emotions in food choice and liking. *Food Research International* 76, Part 2**,** 216-223.

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

Hajcak, G., Dunning, J.P., and Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clinical Neurophysiology* 120**,** 505-510.

Haralick, R.M., Shanmugam, K., and Dinstein, I. (1976). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3**,** 610-621.

Hare, R., Wood, K., Britain, S., and Shadman, J. (1970). Autonomic responses to affective visual stimulation. *Psychophysiology* 7**,** 408-417.

Harmon-Jones, E., and Gable, P.A. (2009). Neural activity underlying the effect of approach-motivated positive affect on narrowed attention. *Psychological Science* 20**,** 406-409.

Harmon-Jones, E., Gable, P.A., and Peterson, C.K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology* 84**,** 451-462.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International journal of cross cultural management* 6**,** 243-266.

Hasenbeck, A., Cho, S., Meullenet, J.F., Tokar, T., Yang, F., Huddleston, E.A., and Seo, H.S. (2014). Color and illuminance level of lighting can modulate willingness to eat bell peppers. *Journal of the Science of Food and Agriculture* 94**,** 2049-2056.

Hasler, D., and Süsstrunk, S.E. (2003). "Measuring colorfulness in natural images", in: *Human Vision and Electronic Imaging VIII.* (eds.) B.E. Rogowitz & N.P. Thrasyvoulos. (Bellingham, WA: The International Society for Optical Engineering).

Hassan, M.A., Malik, A.S., Fofi, D., Saad, N., Karasfi, B., Ali, Y.S., and Meriaudeau, F. (2017). Heart rate estimation using facial video: A review. *Biomedical Signal Processing and Control* 38**,** 346-360.

Hayashi, E.C.S., Gutiérrez Posada, J.E., Maike, V.R.M.L., and Baranauskas, M.C.C. (2016). "Exploring new formats of the Self-Assessment Manikin in the design with children", in: *15th Brazilian Symposium on Human Factors in Computer Systems.* ACM).

Hayn-Leichsenring, G.U., Lehmann, T., and Redies, C. (2017). Subjective ratings of beauty and aesthetics: Correlations with statistical image properties in western oil paintings. *i-Perception* 8**,** 2041669517715474.

He, W., Boesveldt, S., De Graaf, C., and De Wijk, R.A. (2016a). The relation between continuous and discrete emotional responses to food odors with facial expressions and non-verbal reports. *Food Quality and Preference* 48, Part A**,** 130-137.

He, W., De Wijk, R.A., De Graaf, C., and Boesveldt, S. (2016b). Implicit and explicit measurements of affective responses to food odors. *Chemical Senses* 41**,** 661-668.

Hebert, K.R., Valle-Inc, and Hackley, S.A. (2015). Modulation of eyeblink and postauricular reflexes during the anticipation and viewing of food images. *Psychophysiology* 52**,** 509-517.

Herman, C.P. (1993). "Effects of heat on appetite", in: *Nutritional needs in hot environments: applications for military personnel in field operations.* (ed.) B.M. Marriott. National Academy Press).

Herz, R.S., Beland, S.L., and Hellerstein, M. (2004). Changing odor hedonic perception through emotional associations in humans. *International Journal of Comparative Psychology* 17.

Hoefling, A., Likowski, K.U., Deutsch, R., Häfner, M., Seibt, B., Mühlberger, A., Weyers, P., and Strack, F. (2009). When hunger finds no fault with moldy corn: Food deprivation reduces food-related disgust. *Emotion* 9**,** 50-58.

Hoffman, L.D., and Polich, J. (1998). EEG, ERPs and food consumption. *Biological psychology* 48**,** 139-151.

Hogervorst, M.A., Brouwer, A.-M., and Van Erp, J.B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience* 8**,** 322.

Hoogeveen, H.R., Dalenberg, J.R., Renken, R.J., Ter Horst, G.J., and Lorist, M.M. (2015). Neural processing of basic tastes in healthy young and older adults – an fMRI study. *Neuroimage* 119**,** 1-12.

Huisman, G., Bruijnes, M., and Heylen, D.K.J. (2016). "A moving feast: Effects of color, shape and animation on taste associations and taste perceptions", in: *13th International Conference on Advances in Computer Entertainment Technology.* (Osaka, Japan: ACM).

Huisman, G., and Van Hout, M. (2008). "The development of a graphical emotion measurement instrument using caricatured expressions: the LEMtool", in: *Workshop Proceedings Emotion in HCI – Designing for People.* (eds.) C. Peter, E. Crane, M. Fabri, H. Agius & L. Axelrod.).

Huisman, G., Van Hout, M., Van Dijk, E., Van Der Geest, T., and Heylen, D. (2013). "LEMtool: measuring emotions in visual interfaces", in: *The SIGCHI Conference on Human Factors in Computing Systems.* ACM).

Jaeger, S.R., and Ares, G. (2017). Dominant meanings of facial emoji: Insights from Chinese consumers and comparison with meanings from internet resources. *Food Quality and Preference* 62**,** 275-283.

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

Jaeger, S.R., Jørgensen, A.S., Aaslyng, M.D., and Bredie, W.L. (2008). Best–worst scaling: An introduction and initial comparison with monadic rating for preference elicitation with food products. *Food Quality and Preference* 19**,** 579-588.

Jaeger, S.R., Lee, S.M., Kim, K.-O., Chheang, S.L., Jin, D., and Ares, G. (2017a). Measurement of product emotions using emoji surveys: Case studies with tasted foods and beverages. *Food Quality and Preference* 62**,** 46-59.

Jaeger, S.R., Roigard, C.M., and Ares, G. (2018a). Measuring consumers' product associations with emoji and emotion word questionnaires: case studies with tasted foods and written stimuli. *Food Research International* 111**,** 732-747.

Jaeger, S.R., Spinelli, S., Ares, G., and Monteleone, E. (2018b). Linking product-elicited emotional associations and sensory perceptions through a circumplex model based on valence and arousal: Five consumer studies. *Food Research International* 109.

Jaeger, S.R., Vidal, L., Kam, K., and Ares, G. (2017b). Can emoji be used as a direct method to measure emotional associations to food names? Preliminary investigations with consumers in USA and China. *Food quality and preference* 56**,** 38-48.

Jaeger, S.R., Vidal, L., Kam, K., and Ares, G. (2017c). Can emoji be used as a direct method to measure emotional associations to food names? Preliminary investigations with consumers in USA and China. *Food Quality and Preference* 56, Part A**,** 38-48.

Jaeger, S.R., Xia, Y., Lee, P.-Y., Hunter, D.C., Beresford, M.K., and Ares, G. (2018c). Emoji questionnaires can be used with a range of population segments: Findings relating to age, gender and frequency of emoji/emoticon use. *Food Quality and Preference* 68**,** 397-410.

Jager, G., Schlich, P., Tijssen, I., Yao, J., Visalli, M., De Graaf, C., and Stieger, M. (2014). Temporal dominance of emotions: Measuring dynamics of food-related emotions during consumption. *Food Quality and Preference* 37**,** 87-99.

Jensen, C.D., Duraccio, K.M., Barnett, K.A., and Stevens, K.S. (2016). Appropriateness of the food-pics image database for experimental eating and appetite research with adolescents. *Eating Behaviors* 23**,** 195-199.

Johnson, T., Kulesa, P., Cho, Y.I., and Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-cultural psychology* 36**,** 264-277.

Kaneko, D., Hogervorst, M., Toet, A., Van Erp, J.B., Kallen, V., and Brouwer, A.-M. (2019a). Explicit and implicit responses to tasting drinks associated with different tasting experiences. *Sensors* 19**,** 4397.

Kaneko, D., Toet, A., Brouwer, A.-M., Kallen, V., and Van Erp, J.B. (2018a). Methods for evaluating emotions evoked by food experiences: a literature review. *Frontiers in psychology* 9**,** 911.

Kaneko, D., Toet, A., Brouwer, A.-M., Kallen, V., and Van Erp, J.B.F. (2018b). Methods for evaluating emotions evoked by food experiences: A literature review. *Frontiers in Psychology* 9**,** 1-20.

Kaneko, D., Toet, A., Ushiama, S., Brouwer, A.-M., Kallen, V., and Van Erp, J.B. (2019b). EmojiGrid: a 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food. *Food research international* 115**,** 541-551.

Kaneko, D., Toet, A., Ushiama, S., Brouwer, A., Kallen, V., and Van Erp, J. (2019c). EmojiGrid: A 2D pictorial scale for cross-cultural emotion assessment of negatively and positivey valenced food: poster en abstract.

Kaneko, D., Toet, A., Ushiama, S., Brouwer, A.M., Kallen, V., and Van Erp, J.B.F. (2018c). EmojiGrid: a 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food. *Food Research International* In press.

Kawano, Y., and Yanai, K. (2015). "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation", in: *Computer Vision - ECCV 2014 Workshops.* (eds.) L. Agapito, M.M. Bronstein & C. Rother. Springer International Publishing).

Kaye, L.K., Malone, S.A., and Wall, H.J. (2017). Emojis: Insights, affordances, and possibilities for psychological science. *Trends in Cognitive Sciences* 21**,** 66-68.

Kelly, G. (2003). *The psychology of personal constructs: Volume two: Clinical diagnosis and psychotherapy.* Routledge.

Kemps, E., Tiggemann, M., Martin, R., and Elliott, M. (2013). Implicit approach–avoidance associations for craved food cues. *Journal of Experimental Psychology: Applied* 19**,** 30.

Kenney, E., and Adhikari, K. (2016a). Recent developments in identifying and quantifying emotions during food consumption. *Journal of the Science of Food and Agriculture* 96**,** 3627-3630.

Kenney, E., and Adhikari, K. (2016b). Recent developments in identifying and quantifying emotions during food consumption. *Journal of the Science of Food and Agriculture***,** n/a-n/a.

Killgore, W.D.S., Young, A.D., Femia, L.A., Bogorodzki, P., Rogowska, J., and Yurgelun-Todd, D.A. (2003). Cortical and limbic activation during viewing of high- versus low-calorie foods. *NeuroImage* 19**,** 1381-1394.

Kim, I.-A., Kim, M.-A., Van De Velden, M., and Lee, H.-S. (2013). Psychological positioning of bottled tea products: A comparison between two Kansei profiling techniques. *Food Science and Biotechnology* 22**,** 257-268.

Kim, K.O., and O'mahony, M. (1998). A new approach to category scales of intensity I: Traditional versus rank-rating. *Journal of Sensory Studies* 13**,** 241-249.

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

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

Kleinginna, P.R., and Kleinginna, A.M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion* 5**,** 345-379.

Knaapila, A., Laaksonen, O., Virtanen, M., Yang, B., Lagström, H., and Sandell, M. (2017). Pleasantness, familiarity, and identification of spice odors are interrelated and enhanced by consumption of herbs and food neophilia. *Appetite* 109**,** 190-200.

König, L.M., and Renner, B. (2018). Colourful =healthy? Exploring meal colour variety and its relation to food consumption. *Food Quality and Preference* 64**,** 66-71.

Koo, T.K., and Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15**,** 155-163.

Köster, E., and Mojet, J. (2006). Theories of food choice development. *Understanding consumers of food products***,** 93-124.

Köster, E.P. (2003). The psychology of food choice: some often encountered fallacies. *Food Quality and Preference* 14**,** 359-373.

Köster, E.P. (2009). Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference* 20**,** 70-82.

Köster, E.P., and Mojet, J. (2007). "Boredom and the reasons why some new products fail", in: *Consumer—Led Food Product Development.* (ed.) Macfie, H.: Woodhead).

Köster, E.P., and Mojet, J. (2015a). 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.

Köster, E.P., and Mojet, J. (2015b). 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.

Kostyra, E., Rambuszek, M., Waszkiewicz-Robak, B., Laskowski, W., Blicharski, T., and Poławska, E. (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.

Kralj Novak, P., Smailović, J., Sluban, B., and Mozetič, I. (2015). Sentiment of emojis. *PLOS ONE* 10**,** e0144296.

Kranjec, J., Beguš, S., Geršak, G., and Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. *Biomedical signal processing and control* 13**,** 102-112.

Kreibig, S.D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology* 84**,** 394-421.

Kremer, S., Bult, J.H.F., Mojet, J., and Kroeze, J.H.A. (2007). Food perception with age and its relationship to pleasantness. *Chemical Senses* 32**,** 591-602.

Krieglmeyer, R., Deutsch, R., De Houwer, J., and De Raedt, R. (2010). Being moved: Valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science* 21**,** 607-613.

Kuenzel, J., Blanchette, I., Lion, R., Zandstra, E.H., Thomas, A., and El-Deredy, W. (2011). Conditioning specific positive states to unfamiliar flavours influences flavour liking. *Food Quality and Preference* 22**,** 397-403.

Kuenzel, J., Zandstra, E.H., Lion, R., Blanchette, I., Thomas, A., and El-Deredy, W. (2010). Conditioning unfamiliar and familiar flavours to specific positive emotions. *Food Quality and Preference* 21**,** 1105-1107.

Kuijsters, A., Redi, J., De Ruyter, B., Seuntiëns, P., and Heynderickx, I. (2015). Affective ambiences created with lighting for older people. *Lighting Research and Technology* 47**,** 859-875.

Kuppens, P., Tuerlinckx, F., Russell, J.A., and Barrett, L.F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin* 139**,** 917-940.

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K.J., and Russell, J.A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of Personality* 85**,** 530-542.

Kytö, E., Bult, H., Aarts, E., Wegman, J., Ruijschop, R.M., and Mustonen, S. (2019). Comparison of explicit vs. implicit measurements in predicting food purchases. *Food Quality and Preference* 78**,** 103733.

Lagast, S., Gellynck, X., Schouteten, J., De Herdt, V., and De Steur, H. (2017). Consumers' emotions elicited by food: A systematic review of explicit and implicit methods. *Trends in food science & technology* 69**,** 172-189.

Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist* 50**,** 372.

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (2005). "International affective picture system (IAPS): Instruction manual and affective ratings". (Gainesville, Florida: Technical Report A-6, The Center for Research in Psychophysiology, University of Florida).

Lang, P.J., Bradley, M.M., Fitzsimmons, J.R., Cuthbert, B.N., Scott, J.D., Moulder, B., and Nangia, V. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology* 35**,** 199-210.

Lang, P.J., Greenwald, M.K., Bradley, M.M., and Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30**,** 261-273.

Lange, C., Martin, C., Chabanet, C., Combris, P., and Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for Champagne: comparison with hedonic scores. *Food Quality and Preference* 13**,** 597-608.

Larson, C., Aronoff, J., and Steuer, E. (2012). Simple geometric shapes are implicitly associated with affective value. *Motivation and Emotion* 36**,** 404-413.

Laurans, G.F.G., and Desmet, P.M.A. (2012a). "Introducing PrEmo2: New directions for the non-verbal measurement of emotion in design", in: *8th International Conference on Design and Emotion.* (eds.) J. Brassett, P. Hekkert, G. Ludden, M. Malpass & J. Mcdonnell. (London, UK: Central Saint Martins College of Arts & Design).

Laurans, G.F.G., and Desmet, P.M.A. (2012b). "Introducing PrEmo2: New directions for the non-verbal measurement of emotion in design", in: *Proceedings of the 8th International Conference on Design and Emotion.* (London, UK).

Lawless, H.T., and Heymann, H. (2010). *Sensory evaluation of food: principles and practices.* Springer Science & Business Media.

Lee, J.W., Jones, P.S., Mineyama, Y., and Zhang, X.E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health* 25**,** 295-306.

Levitsky, D.A., Iyer, S., and Pacanowski, C.R. (2012). Number of foods available at a meal determines the amount consumed. *Eating Behaviors* 13**,** 183-187.

Lévy, C.M., and Köster, E.P. (1999). The relevance of initial hedonic judgements in the prediction of subtle food choices. *Food Quality and Preference* 10**,** 185-200.

Libby Jr, W.L., Lacey, B.C., and Lacey, J.I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology* 10**,** 270-294.

Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food quality and preference* 22**,** 733-747.

Lin, I.Y. (2004). Evaluating a servicescape: the effect of cognition and emotion. *International Journal of Hospitality Management* 23**,** 163-178.

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., and Barrett, L.F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences* 35**,** 121-143.

Liu, R., Hannum, M., and Simons, C.T. (2018). Using immersive technologies to explore the effects of congruent and incongruent contextual cues on context recall, product evaluation time, and preference and liking during consumer hedonic testing. *Food Research International* in press.

Locher, J.L., Yoels, W.C., Maurer, D., and Van Ells, J. (2005). Comfort foods: an exploratory journey into the social and emotional significance of food. *Food & Foodways* 13**,** 273-297.

Lopatovska, I., and Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing & Management* 47**,** 575-592.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering* 4**,** R1.

Lottridge, D., Chignell, M., and Yasumura, M. (2011). Identifying emotion through implicit and explicit measures: cultural differences, cognitive load, and immersion. *IEEE Transactions on Affective Computing* 3**,** 199-210.

Lottridge, D., Chignell, M., and Yasumura, M. (2012). Identifying emotion through implicit and explicit measures: Cultural differences, cognitive load, and immersion. *IEEE Transactions on Affective Computing* 3**,** 199-210.

Lucassen, M.P., Gevers, T., and Gijsenij, A. (2011). Texture affects color emotion. *Color Research & Application* 36**,** 426-436.

Lundqvist, D., Esteves, F., and Öhman, A. (2004). The face of wrath: The role of features and configurations in conveying social threat. *Cognition & Emotion* 18**,** 161-182.

Macht, M. (1999). Characteristics of eating in anger, fear, sadness and joy. *Appetite* 33**,** 129-139.

Macht, M. (2008). How emotions affect eating: a five-way model. *Appetite* 50**,** 1-11.

Macht, M., Roth, S., and Ellgring, H. (2002). Chocolate eating in healthy men during experimentally induced sadness and joy. *Appetite* 39**,** 147-158.

Macht, M., and Simons, G. (2000). Emotions and eating in everyday life. *Appetite* 35**,** 65-71.

Madan, C.R., Bayer, J., Gamer, M., Lonsdorf, T.B., and Sommer, T. (2018). Visual complexity and affect: Ratings reflect more than meets the eye. *Frontiers in Psychology* 8.

Madaniyazi, L., Zhou, Y., Li, S., Williams, G., Jaakkola, J.J., Liang, X., Liu, Y., Wu, S., and Guo, Y. (2016). Outdoor temperature, heart rate and blood pressure in Chinese adults: effect modification by individual characteristics. *Scientific reports* 6**,** 1-9.

Majima, Y., Nishiyama, K., Nishihara, A., and Hata, R. (2017). Conducting online behavioral research using crowdsourcing services in Japan. *Frontiers in Psychology* 8.

Manzocco, L., Rumignani, A., and Lagazio, C. (2013). Emotional response to fruit salads with different visual quality. *Food Quality and Preference* 28**,** 17-22.

Marchewka, A., Żurawski, Ł., Jednoróg, K., and Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods* 46**,** 596-610.

Marczinski, C.A., Stamates, A.L., Ossege, J., Maloney, S.F., Bardgett, M.E., and Brown, C.J. (2014). Subjective state, blood pressure, and behavioral control changes produced by an "Energy Shot". *J Caffeine Res* 4**,** 57-63.

Marin, M.M., and Leder, H. (2013). Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PLOS ONE* 8**,** e72412.

Matsuda, Y., Hoashi, H., and Yanai, K. (2012). "Recognition of multiple-food images by detecting candidate regions", in: *2012 IEEE International Conference on Multimedia and Expo.*).

Mattek, A.M., Wolford, G.L., and Whalen, P.J. (2017). A mathematical model captures the structure of subjective affect. *Perspectives on Psychological Science* 12**,** 508-526.

Mauss, I.B., and Robinson, M.D. (2009). Measures of emotion: A review. *Cognition and Emotion* 23**,** 209-237.

Mccarron, A., and Tierney, K.J. (1989). The effect of auditory stimulation on the consumption of soft drinks. *Appetite* 13**,** 155-159.

Mcgaugh, J.L. (2006). Make mild moments memorable: add a little arousal. *Trends in cognitive sciences* 10**,** 345-347.

Mcnair, D., Lorr, M., and Droppleman, L. (1971). *Mannual for the Profile of Mood States (POMS).* San Diego, CA: Educational and Industrial Testing Service.

Mehrabian, A., and Russell, J.A. (1974a). *An approach to environmental psychology.* Cambridge, MA, US: The MIT Press.

Mehrabian, A., and Russell, J.A. (1974b). *An approach to environmental psychology.* Boston, MA, USA: The MIT Press.

Meiselman, H.L. (2013a). The future in sensory/consumer research: ………......evolving to a better science. *Food Quality and Preference* 27**,** 208-214.

Meiselman, H.L. (2013b). The future in sensory/consumer research: Evolving to a better science. *Food Quality and Preference* 27**,** 208-214.

Meiselman, H.L. (2015a). A review of the current state of emotion research in product development. *Food Research International* 76**,** 192-199.

Meiselman, H.L. (2015b). A review of the current state of emotion research in product development. *Food Research International* 76, Part 2**,** 192-199.

Meiselman, H.L. (2021). The (gradual) development of emotion measurement for food. *Current Opinion in Food Science*.

Meléndez-Martínez, A.J., Vicario, I.M., and Heredia, F.J. (2005). Correlation between visual and instrumental colour measurements of orange juice dilutions: effect of the background. *Food Quality and Preference* 16**,** 471-478.

Mendes, W.B. (2009). Assessing autonomic nervous system activity. *Methods in social neuroscience* 118**,** 147.

Meule, A., Lender, A., Richard, A., Dinic, R., and Blechert, J. (2019). Approach–avoidance tendencies towards food: Measurement on a touchscreen and the role of attention and food craving. *Appetite* 137**,** 145-151.

Meule, A., Richard, A., Lender, A., Dinic, R., Brockmeyer, T., Rinck, M., and Blechert, J. (2020). Measuring approach–avoidance tendencies towards food with touchscreen-based arm movements. *Psychological research* 84**,** 1789-1800.

Miccoli, L., Delgado, R., Guerra, P., Versace, F., Rodríguez-Ruiz, S., and Fernández-Santaella, M.C. (2016). Affective pictures and the Open Library of Affective Foods (OLAF): tools to investigate emotions toward food in adults. *PLOS ONE* 11**,** e0158991.

Miccoli, L., Delgado, R., Rodríguez-Ruiz, S., Guerra, P., García-Mármol, E., and Fernández-Santaella, M.C. (2014). Meet OLAF, a good friend of the IAPS! The Open Library of Affective Foods: A tool to investigate the emotional impact of food in adolescents. *PLoS ONE* 9**,** e114515.

Michel, C., Velasco, C., Fraemohs, P., and Spence, C. (2015a). Studying the impact of plating on ratings of the food served in a naturalistic dining context. *Appetite* 90**,** 45-50.

Michel, C., Velasco, C., and Spence, C. (2015b). Cutlery matters: heavy cutlery enhances diners' enjoyment of the food served in a realistic dining environment. *Flavour* 4**,** 26.

Mielby, L.H., Kildegaard, H., Gabrielsen, G., Edelenbos, M., and Thybo, A.K. (2012). Adolescent and adult visual preferences for pictures of fruit and vegetable mixes – Effect of complexity. *Food Quality and Preference* 26**,** 188-195.

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., and Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying Interpretations of Emoji", in: *Tenth International AAAI Conference on Web and Social Media (ICWSM 2016).* AAAI).

Mojet, J., Dürrschmid, K., Danner, L., Jöchl, M., Heiniö, R.L., Holthuysen, N., and Köster, E. (2015). Are implicit emotion measurements evoked by food unrelated to liking? *Food Research International* 76, Part 2**,** 224-232.

Moore, A., Steiner, C.M., and Conlan, O. (2013). "Design and development of an empirical smiley-based affective instrument", in: *UMAP 2013 Extended Proceedings. EMPIRE 2013: Emotions and Personality in Personalized Services.* (ed.) S.J. Alfredo. (Rome, Italy: CEUR Workshop Proceedings).

Moore, D.J., and Konrath, S. (2015). "I can almost taste it:" Why people with strong positive emotions experience higher levels of food craving, salivation and eating intentions. *Journal of Consumer Psychology* 25**,** 42-59.

Muñoz, M.Á., Viedma-Del-Jesus, M.I., Fernández-Santaella, M.C., Peralta-Ramírez, M.I., Cepeda-Benito, A., and Vila, J. (2010). Assessment of tobacco craving by means of the affective image visualization paradigm. *Motivation and Emotion* 34**,** 93-103.

Nagel, F., Kopiez, R., Grewe, O., and Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* 39**,** 283-290.

Narumi, T., Ban, Y., Kajinami, T., Tanikawa, T., and Hirose, M. (Year). "Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality", in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM), 109-118.

Narumi, T., Ban, Y., Kajinami, T., Tanikawa, T., and Hirose, M. (2012b). "Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality", in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* (Austin, Texas, USA: ACM).

Narumi, T., Nishizaka, S., Kajinami, T., Tanikawa, T., and Hirose, M. (Year). "Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction", in: *Proceedings of the SIGCHI conference on human factors in computing systems*: ACM), 93-102.

Nestrud, M.A., Meiselman, H.L., King, S.C., Lesher, L.L., and Cardello, A.V. (2013). "A method for development of EsSense25–A shorter version of the EsSense (R) profile", in: *10th Pangborn Sensory Science Symposium 2013: Rio de Janeiro, Brazil.* (Rio de Janeiro, Brazil).

Nestrud, M.A., Meiselman, H.L., King, S.C., Lesher, L.L., and Cardello, A.V. (2016a). Development of EsSense25, a shorter version of the EsSense Profile. *Food Quality and Preference* 48**,** 107-117.

Nestrud, M.A., Meiselman, H.L., King, S.C., Lesher, L.L., and Cardello, A.V. (2016b). Development of EsSense25, a shorter version of the EsSense Profile®. *Food Quality and Preference* 48**,** 107-117.

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

Nisbett, R.E., Peng, K., Choi, I., and Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review* 108**,** 291-310.

Nishimura, Y. (2015). "A sociolinguistic analysis of emoticon usage in Japanese blogs: Variation by age, gender, and topic", in: *16th Annual Meeting of the Association of Internet Researchers.* (Phoenix, AZ, USA: AoIR).

Nishizawa, M., Jiang, W., and Okajima, K. (2016). "Projective-AR system for customizing the appearance and taste of food", in: *The 2016 workshop on Multimodal Virtual and Augmented Reality.* (Tokyo, Japan: ACM).

Nordbo, K., Milne, D., Calvo, R.A., and Allman-Farinelli, M. (Year). "Virtual Food Court: A VR environment to Assess People's Food Choices", in: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*: ACM), 69-72.

Nordbo, K., Milne, D., Calvo, R.A., and Allman-Farinelli, M. (2015b). "Virtual food court: A VR environment to assess people's food choices", in: *Annual Meeting of the Australian Special Interest Group for Computer Human Interaction.* (Parkville, VIC, Australia: ACM).

Obaid, M., Dünser, A., Moltchanova, E., Cummings, D., Wagner, J., and Bartneck, C. (2015). "LEGO Pictorial Scales for Assessing Affective Response", in: *Human-Computer Interaction – INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part I.* (eds.) J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque & M. Winckler. (Cham: Springer International Publishing).

Oberfeld, D., Hecht, H., Allendorf, U., and Wickelmaier, F. (2009). Ambient lighting modifies the flavor of wine. *Journal of Sensory Studies* 24**,** 797-832.

Obrist, M., Gatti, E., Maggioni, E., Vi, C.T., and Velasco, C. (2017). Multisensory experiences in HCI. *IEEE MultiMedia* 24**,** 9-13.

Obrist, M., Velasco, C., Vi, C., Ranasinghe, N., Israr, A., Cheok, A., Spence, C., and Gopalakrishnakone, P. (2016). Sensing the future of HCI: touch, taste, and smell user interfaces. *Interactions* 23**,** 40-49.

Ohgami, Y., Kotani, Y., Tsukamoto, T., Omura, K., Inoue, Y., Aihara, Y., and Nakayama, M. (2006). Effects of monetary reward and punishment on stimulus-preceding negativity. *Psychophysiology* 43**,** 227-236.

Oliver, G., and Wardle, J. (1999). Perceived effects of stress on food choice. *Physiology & behavior* 66**,** 511-515.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* 2011**,** 1.

Pallavicini, F., Serino, S., Cipresso, P., Pedroli, E., Chicchi Giglioli, I.A., Chirico, A., Manzoni, G.M., Castelnuovo, G., Molinari, E., and Riva, G. (2016). Testing augmented reality for cue exposure in obese patients: An exploratory study. *Cyberpsychology, Behavior, and Social Networking* 19**,** 107-114.

Palus, H. (2005). "Colourfulness of the image and its application in image filtering", in: *Fifth IEEE International Symposium on Signal Processing and Information Technology.* (Washington,USA: IEEE Press).

Pan, J., and Tompkins, W.J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering***,** 230-236.

Papousek, I., Weiss, E.M., Schulter, G., Fink, A., Reiser, E.M., and Lackner, H.K. (2014). Prefrontal EEG alpha asymmetry changes while observing disaster happening to other people: cardiac correlates and prediction of emotional impact. *Biological psychology* 103**,** 184-194.

Pashler, H., and Wagenmakers, E.J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science:A Crisis of Confidence? *Perspectives on Psychological Science* 7**,** 528-530.

Patel, S.H., and Azzam, P.N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International journal of medical sciences* 2**,** 147.

Pfurtscheller, G., Stancak Jr, A., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. *International journal of psychophysiology* 24**,** 39-46.

Pineau, N., Cordelle, S., and Schlich, P. (Year). "Temporal dominance of sensations: A new technique to record several sensory attributes simultaneously over time", in: *5th Pangborn symposium*), 121.

Piqueras-Fiszman, B., Alcaide, J., Roura, E., and Spence, C. (2012). Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it. *Food Quality and Preference* 24**,** 205-208.

Piqueras-Fiszman, B., Giboreau, A., and Spence, C. (2013). Assessing the influence of the color of the plate on the perception of a complex food in a restaurant setting. *Flavour* 2**,** 24.

Piqueras-Fiszman, B., Kraus, A.A., and Spence, C. (2014). “Yummy” versus “Yucky”! Explicit and implicit approach–avoidance motivations towards appealing and disgusting foods. *Appetite* 78**,** 193-202.

Piqueras-Fiszman, B., and Spence, C. (2014). Colour, pleasantness, and consumption behaviour within a meal. *Appetite* 75**,** 165-172.

Poole, N.D., Martı, L., and Giménez, F.V. (2007). Quality perceptions under evolving information conditions: Implications for diet, health and consumer satisfaction. *Food Policy* 32**,** 175-188.

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

Prada, M., Rodrigues, D., Garrido, M.V., and Lopes, J. (2017). Food-pics-PT: Portuguese validation of food images in 10 subjective evaluative dimensions. *Food Quality and Preference* 61**,** 15-25.

Prescott, J. (2017). Some considerations in the measurement of emotions in sensory and consumer research. *Food quality and preference* 62**,** 360-368.

Proulx, R., and Parrott, L. (2008). Measures of structural complexity in digital images for monitoring the ecological signature of an old-growth forest ecosystem. *Ecological Indicators* 8**,** 270-284.

Quirin, M., Kazén, M., Rohrmann, S., and Kuhl, J. (2009). Implicit but Not Explicit Affectivity Predicts Circadian and Reactive Cortisol: Using the Implicit Positive and Negative Affect Test. *Journal of Personality* 77**,** 401-426.

Raudenbush, B., and Frank, R.A. (1999). Assessing food neophobia: The role of stimulus familiarity. *Appetite* 32**,** 261-271.

Redies, C., Amirshahi, S.A., Koch, M., and Denzler, J. (2012). "PHOG-Derived aesthetic measures applied to color photographs of artworks, natural scenes and objects," in *Computer Vision – ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7-13, 2012, Proceedings, Part I,* eds. A. Fusiello, V. Murino & R. Cucchiara. (Berlin, Heidelberg: Springer Berlin Heidelberg), 522-531.

Reinoso Carvalho, F., Van Ee, R., and Rychtarikova, M. (2016). "Matching soundscapes and music with food types", in: *Proceedings of Euroregio 2016.* Sociedade Portuguesa de Acústica).

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology* 67**,** 525-539.

Rellecke, J., Palazova, M., Sommer, W., and Schacht, A. (2011). On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain and Cognition* 77**,** 23-32.

Ribeiro, R.L., Pompéia, S., and Bueno, O.F.A. (2005). Comparison of Brazilian and American norms for the international affective picture system (IAPS). *Revista Brasileira de Psiquiatria* 27**,** 208-215.

Richins, M.L. (1997). Measuring emotions in the consumption experience. *Journal of consumer research* 24**,** 127-146.

Riegel, M., Moslehi, A., Michałowski, J.M., Żurawski, Ł., Horvat, M., Wypych, M., Jednoróg, K., and Marchewka, A. (2017). Nencki affective picture system: Cross-cultural study in Europe and Iran. *Frontiers in psychology* 8**,** 274.

Rimm-Kaufman, S.E., and Kagan, J. (1996). The psychological significance of changes in skin temperature. *Motivation and Emotion* 20**,** 63-78.

Roggeveen, A.L., Grewal, D., Townsend, C., and Krishnan, R. (2015). The impact of dynamic presentation format on consumer preferences for hedonic products and services. *Journal of Marketing* 79**,** 34-49.

Rolls, B.J., Rowe, E.A., and Rolls, E.T. (1982). How sensory properties of foods affect human feeding behavior. *Physiology & Behavior* 29**,** 409-417.

Rolls, E.T., and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology* 86**,** 216-244.

Rosas-Nexticapa, M., Angulo, O., and O'mahony, M. (2005). How well does the 9-point hedonic scale predict purchase frequency? *Journal of Sensory Studies* 20**,** 313-331.

Rosenholtz, R., Li, Y., and Nakano, T. (2007). Measuring visual clutter. *Journal of Vision* 7**,** 1-22.

Roth, W.T. (1983). "8 A Comparison of P300 and Skin Conductance Response," in *Advances in Psychology*. Elsevier), 177-199.

Rousmans, S., Robin, O., Dittmar, A., and Vernet-Maury, E. (2000). Autonomic nervous system responses associated with primary tastes. *Chemical Senses* 25**,** 709-718.

Rozin, P. (1988). "Cultural approaches to human food preferences", in: *Nutritional modulation of neural function.* (eds.) J.E. Morley, M.B. Sterman & J.H. Walsh. Academic Press).

Rozin, P. (1996). Towards a psychology of food and eating: From motivation to module to model to marker, morality, meaning, and metaphor. *Current Directions in Psychological Science* 5**,** 18-24.

Rozin, P. (2006). The integration of biological, social, cultural and psychological influences on food choice. *Frontiers in nutritional science* 3**,** 19.

Runge, N., Hellmeier, M., Wenig, D., and Malaka, R. (2016). "Tag your emotions: a novel mobile user interface for annotating images with emotions", in: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct.* (Florence, Italy: ACM).

Russell, C.G., Busson, A., Flight, I., Bryan, J., Van Lawick Van Pabst, J.A., and Cox, D.N. (2004). A comparison of three laddering techniques applied to an example of a complex food choice. *Food Quality and Preference* 15**,** 569-583.

Russell, J.A. (1980a). A circumplex model of affect. *Journal of personality and social psychology* 39**,** 1161.

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

Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological review* 110**,** 145.

Russell, J.A., and Barrett, L.F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology* 76**,** 805.

Russell, J.A., and Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of personality and social psychology* 38**,** 311.

Russell, J.A., Weiss, A., and Mendelson, G.A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology* 57**,** 493-502.

Saeed, A., Trajanovski, S., Van Keulen, M., and Van Erp, J. (Year). "Deep physiological arousal detection in a driving simulator using wearable sensors", in: *IEEE International Conference on Data Miningworkshop: Data Mining in Biomedical Informatics and Healthcare (DMBIH)*).

Said, C.P., Sebe, N., and Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion* 9**,** 260-264.

Salvy, S.-J., Kieffer, E., and Epstein, L.H. (2008). Effects of social context on overweight and normal-weight children's food selection. *Eating behaviors* 9**,** 190-196.

Samant, S.S., Chapko, M.J., and Seo, H.-S. (2017). Predicting consumer liking and preference based on emotional responses and sensory perception: A study with basic taste solutions. *Food research international* 100**,** 325-334.

Samant, S.S., and Seo, H.-S. (2018). Using both emotional responses and sensory attribute intensities to predict consumer liking and preference toward vegetable juice products. *Food Quality and Preference* In press.

Sánchez, J.A., Hernández, N.P., Penagos, J.C., and Ostróvskaya, Y. (2006). "Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states", in: *VII Brazilian symposium on Human factors in computing systems.* (Natal, RN, Brazil: ACM).

Sander, D. (2013). Models of emotion: The affective neuroscience approach.

Sander, D., Grandjean, D., and Scherer, K.R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural networks* 18**,** 317-352.

Sandt, A.R., Sloan, D.M., and Johnson, K.J. (2009). Measuring appetitive responding with the postauricular reflex. *Psychophysiology* 46**,** 491-497.

Schacht, A., and Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition* 69**,** 538-550.

Scherer, K.R. (2005). What are emotions? And how can they be measured? *Social Science Information* 44**,** 695-729.

Schifferstein, H.N. (2015). Employing consumer research for creating new and engaging food experiences in a changing world. *Current Opinion in Food Science* 3**,** 27-32.

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

Schifferstein, H.N.J., Howell, B.F., and Pont, S.C. (2017). Colored backgrounds affect the attractiveness of fresh produce, but not it's perceived color. *Food Quality and Preference* 56, Part A**,** 173-180.

Schmitz, T.W., and Johnson, S.C. (2007). Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neuroscience & Biobehavioral Reviews* 31**,** 585-596.

Schöning, J., Rogers, Y., and Krüger, A. (2012). Digitally enhanced food. *IEEE Pervasive Computing* 11**,** 4-6.

Schouteten, J.J., De Steur, H., De Pelsmaeker, S., Lagast, S., De Bourdeaudhuij, I., and Gellynck, X. (2015). An integrated method for the emotional conceptualization and sensory characterization of food products: The EmoSensory-« Wheel. *Food Research International* 78**,** 96-107.

Schouteten, J.J., Verwaeren, J., Lagast, S., Gellynck, X., and De Steur, H. (2018). Emoji as a tool for measuring children's emotions when tasting food. *Food Quality and Preference* 68**,** 322-331.

Schreuder, E., Van Erp, J., Toet, A., and Kallen, V.L. (2016). Emotional Responses to Multisensory Environmental Stimuli: A Conceptual Framework and Literature Review. *SAGE Open* 6.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology* 51**,** 154-165.

Schüssler, P., Kluge, M., Yassouridis, A., Dresler, M., Uhr, M., and Steiger, A. (2012). Ghrelin levels increase after pictures showing food. *Obesity* 20**,** 1212-1217.

Schutz, H.G., and Cardello, A.V. (2001). A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *Journal of Sensory Studies* 16**,** 117-159.

Sedghamiz, H. (2014). Matlab implementation of Pan Tompkins ECG QRS detector. *Code Available at the File Exchange Site of MathWorks*.

Shrout, P.E., and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86**,** 420-428.

Siegel, S.F., and Risvik, E. (1987). Cognitive Set and Food Acceptance. *Journal of food science* 1987 v.52 no.3**,** pp. 825-826.

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

Simmons, W.K., Martin, A., and Barsalou, L.W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex* 15**,** 1602-1608.

Singh, M. (2014). Mood, food, and obesity. *Frontiers in psychology* 5**,** 925.

Slater, S., and Yani-De-Soriano, M. (2010). Researching consumers in multicultural societies: Emerging methodological issues. *Journal of Marketing Management* 26**,** 1143-1160.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M.M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39**,** 701-711.

Small, D.M., Zatorre, R.J., Dagher, A., Evans, A.C., and Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolateFrom pleasure to aversion. *Brain* 124**,** 1720-1733.

Sokhadze, E.M. (2007). Effects of music on the recovery of autonomic and electrocortical activity after stress induced by aversive visual stimuli. *Applied psychophysiology and biofeedback* 32**,** 31-50.

Solarz, A.K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of experimental psychology* 59**,** 239.

Soto, J.A., Lee, E.A., and Roberts, N.A. (2016). Convergence in feeling, divergence in physiology: How culture influences the consequences of disgust suppression and amplification among European Americans and Asian Americans. *Psychophysiology* 53**,** 41-51.

Soto, J.A., Levenson, R.W., and Ebling, R. (2005). Cultures of moderation and expression: emotional experience, behavior, and physiology in Chinese Americans and Mexican Americans. *Emotion* 5**,** 154.

Spence, C. (2011a). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics* 73**,** 971-995.

Spence, C. (2011b). Mouth-watering: The influence of environmental and cognitive factors on salivation and gustatory/flavor perception. *Journal of Texture Studies* 42**,** 157-171.

Spence, C. (2015a). Leading the consumer by the nose: on the commercialization of olfactory design for the food and beverage sector. *Flavour* 4**,** 31.

Spence, C. (2015b). On the psychological impact of food colour. *Flavour* 4**,** 21.

Spence, C. (2018). Background colour & its impact on food perception & behaviour. *Food Quality and Preference* 68**,** 156-166.

Spence, C., Levitan, C.A., Shankar, M.U., and Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosensory Perception* 3**,** 68-84.

Spence, C., Okajima, K., Cheok, A.D., Petit, O., and Michel, C. (2015). Eating with our eyes: From visual hunger to digital satiation. *Brain and Cognition* 110**,** 53-63.

Spence, C., and Piqueras-Fiszman, B. (2014). *The perfect meal: the multisensory science of food and dining.* John Wiley & Sons.

Spence, C., and Shankar, M.U. (2010). The influence of auditory cues on the perception of, and responses to, food and drink. *Journal of Sensory Studies* 25**,** 406-430.

Spence, C., Velasco, C., and Knoeferle, K. (2014). A large sample study on the influence of the multisensory environment on the wine drinking experience. *Flavour* 3**,** 1-12.

Sperber, A.D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 126**,** S124-S128.

Spielberger, C.D. (1983). *State-Trait Anxiety Inventory for adults.* Mountain View, California: Consulting Psychologists Press, Inc.

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

Spinelli, S., Masi, C., Zoboli, G.P., Prescott, J., and Monteleone, E. (2015). Emotional responses to branded and unbranded foods. *Food Quality and Preference* 42**,** 1-11.

Steiner, J.E., Glaser, D., Hawilo, M.E., and Berridge, K.C. (2001). Comparative expression of hedonic impact: affective reactions to taste by human infants and other primates. *Neuroscience & Biobehavioral Reviews* 25**,** 53-74.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication* 42**,** 73-93.

Stewart, P.C., and Goss, E. (2013). Plate shape and colour interact to influence taste and quality judgments. *Flavour* 2**,** 27.

Stockburger, J., Renner, B., Weike, A.I., Hamm, A.O., and Schupp, H.T. (2009). Vegetarianism and food perception. Selective visual attention to meat pictures. *Appetite* 52**,** 513-516.

Stroebele, N., and De Castro, J.M. (2004). Effect of ambience on food intake and food choice. *Nutrition* 20**,** 821-838.

Suk, H.-J., and Irtel, H. (2010). Emotional response to color across media. *Color Research & Application* 35**,** 64-77.

Suk, H.-J., Park, G.-L., and Kim, Y. (2012). Bon Appétit! An investigation about the best and worst color combinations of lighting and food. *Journal of Literature and Art Studies* 2**,** 559-566.

Swan, A.B., Cohen, A., Evans, S.R., and Drescher, B.A. (2013). Influence of taste quality on affective state. *Psi Chi Journal of Psychological Research* 18.

Swaney-Stueve, M., Jepsen, T., and Deubler, G. (2018). The emoji scale: A facial scale for the 21st century. *Food Quality and Preference* 68**,** 183-190.

Swindells, C., Maclean, K.E., Booth, K.S., and Meitner, M. (2006). "A case-study of affect measurement tools for physical user interface design", in: *Graphics Interface 2006 (GI '06).* Canadian Human-Computer Communications Society).

Szocs, C., and Lefebvre, S. (2017). Spread or stacked? Vertical versus horizontal food presentation, portion size perceptions, and consumption. *Journal of Business Research* 75**,** 249-257.

Tahir, Y., Rahman, A.U., and Ravana, S.D. (2021). An affect-based classification of emotions associated with images of food. *Journal of Food Measurement and Characterization* 15**,** 519-530.

Takahashi, K., Doman, K., Kawanishi, Y., Hirayama, T., Ide, I., Deguchi, D., and Murase, H. (2017). "Estimation of the attractiveness of food photography focusing on main ingredients", in: *9th Workshop on Multimedia for Cooking and Eating Activities, in conjunction with The 2017 International Joint Conference on Artificial Intelligence.* (Melbourne, Australia: ACM).

Tena, L.M.V. (2019). *The body as a tattletale: Physiological responses to food stimuli perception within the context of expectations.* Wageningen University.

Thammasan, N., Stuldreher, I.V., Schreuders, E., Giletta, M., and Brouwer, A.-M. (2020). A usability study of physiological measurement in school using wearable sensors. *Sensors* 20**,** 5380.

Thompson, E.R. (2007). Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology* 38**,** 227-242.

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

Tigwell, G.W., and Flatla, D.R. (2016). "Oh that's what you meant!: reducing emoji misunderstanding", in: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct.* (Florence, Italy: ACM).

Tipples, J., Atkinson, A.P., and Young, A.W. (2002). The eyebrow frown: A salient social signal. *Emotion* 2**,** 288-296.

Toepel, U., Knebel, J.-F., Hudry, J., Le Coutre, J., and Murray, M.M. (2012). Gender and weight shape brain dynamics during food viewing. *PLOS ONE* 7**,** e36778.

Toet, A. (2016). Natural dynamic backgrounds affect perceived facial dominance. *Matters*.

Toet, A., Bijlsma, M., and Brouwer, A.-M. (2017). Stress response and facial trustworthiness judgments in civilians and military. *Sage open* 7**,** 2158244017725386.

Toet, A., Eijsman, S., Liu, Y., Donker, S., Kaneko, D., Brouwer, A.-M., and Van Erp, J.B. (2019a). The Relation Between Valence and Arousal in Subjective Odor Experience. *Chemosensory Perception***,** 1-11.

Toet, A., Eijsman, S., Liu, Y., Donker, S., Kaneko, D., Brouwer, A.-M., and Van Erp, J.B.F. (2019b). The Relation Between Valence and Arousal in Subjective Odor Experience. *Chemosensory Perception***,** 1-11.

Toet, A., Henselmans, M., Lucassen, M.P., and Gevers, T. (2012). Emotional effects of dynamic textures. *i-Perception* 2**,** 969-991.

Toet, A., Kaneko, D., Kruijf, I.D., Ushiama, S., Schaik, M.V., Brouwer, A.-M., Kallen, V., and Van Erp, J.B. (2019c). CROCUFID: A cross-cultural food image database for research on food elicited affective responses. *Frontiers in psychology* 10**,** 58.

Toet, A., Kaneko, D., Ushiama, S., Hoving, S., De Kruijf, I., Brouwer, A.-M., Kallen, V., and Van Erp, J.B. (2018a). EmojiGrid: A 2D pictorial scale for the assessment of food elicited emotions. *Frontiers in psychology* 9**,** 2396.

Toet, A., Kaneko, D., Ushiama, S., Hoving, S., De Kruijf, I., Brouwer, A.M., Kallen, V., and Van Erp, J.B.F. (2018b). EmojiGrid: a 2D pictorial scale for the assessment of food elicited emotions. *Frontiers in Psychology* In review (preprint available at https://figshare.com/s/5d4ebf70d68399e0205d)

Toet, A., and Van Erp, J.B. (2019). The EmojiGrid as a tool to assess experienced and perceived emotions. *Psych* 1**,** 469-481.

Toet, A., Van Schaik, C.S., Kaneko, D., and Van Erp, J.B.F. (2019d). Do food cinemagraphs evoke stronger appetitive responses than stills? *International Journal of Food Design* 4**,** in press.

Torrico, D.D., Fuentes, S., Viejo, C.G., Ashman, H., and Dunshea, F.R. (2019). Cross-cultural effects of food product familiarity on sensory acceptability and non-invasive physiological responses of consumers. *Food research international* 115**,** 439-450.

Torrico, D.D., Fuentes, S., Viejo, C.G., Ashman, H., Gunaratne, N.M., Gunaratne, T.M., and Dunshea, F.R. (2018a). Images and chocolate stimuli affect physiological and affective responses of consumers: A cross-cultural study. *Food Quality and Preference* 65**,** 60-71.

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

Tsai, J.L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science* 2**,** 242-259.

Tsai, J.L., Chentsova-Dutton, Y., Freire-Bebeau, L., and Przymus, D.E. (2002). Emotional expression and physiology in European Americans and Hmong Americans. *Emotion* 2**,** 380.

Tsai, J.L., Levenson, R.W., and Carstensen, L.L. (2000). Autonomic, subjective, and expressive responses to emotional films in older and younger Chinese Americans and European Americans. *Psychology and aging* 15**,** 684.

Tsourides, K., Shariat, S., Nejati, H., Gandhi, T.K., Cardinaux, A., Simons, C.T., Cheung, N.M., Pavlovic, V., and Sinha, P. (2016). Neural correlates of the food/non-food visual distinction. *Biological Psychology* 115**,** 35-42.

Tu, Y., Yang, Z., and Ma, C. (2016). The taste of plate: How the spiciness of food is affected by the color of the plate used to serve It. *Journal of Sensory Studies* 31**,** 50-60.

Tuorila, H., Lähteenmäki, L., Pohjalainen, L., and Lotti, L. (2001). Food neophobia among the Finns and related responses to familiar and unfamiliar foods. *Food quality and preference* 12**,** 29-37.

Uchida, Y., and Kitayama, S. (2009). Happiness and unhappiness in east and west: themes and variations. *Emotion* 9**,** 441.

Ung, C.-Y., Menozzi, M., Hartmann, C., and Siegrist, M. (2018a). Innovations in consumer research: The virtual food buffet. *Food Quality and Preference* 63**,** 12-17.

Ung, C.-Y., Menozzi, M., Hartmann, C., and Siegrist, M. (2018b). Innovations in consumer research: The virtual food buffet. *Food Quality and Preference* 63**,** 12-17.

Van Beers, J.J., Kaneko, D., Stuldreher, I.V., Zech, H.G., and Brouwer, A.-M. (Year). "An Accessible Tool to Measure Implicit Approach-Avoidance Tendencies Towards Food Outside the Lab", in: *Companion Publication of the 2020 International Conference on Multimodal Interaction*), 307-311.

Van Der Laan, L.N., De Ridder, D.T.D., Viergever, M.A., and Smeets, P.a.M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage* 55**,** 296-303.

Van Ittersum, K., and Wansink, B. (2012). Plate size and color suggestibility: The Delboeuf Illusion's bias on serving and eating behavior. *Journal of Consumer Research* 39**,** 215-228.

Van Zyl, H., and Meiselman, H.L. (2015). The roles of culture and language in designing emotion lists: Comparing the same language in different English and Spanish speaking countries. *Food Quality and Preference* 41**,** 201-213.

Van Zyl, H., and Meiselman, H.L. (2016a). An update on the roles of culture and language in designing emotion lists: English, Spanish and Portuguese. *Food quality and preference* 51**,** 72-76.

Van Zyl, H., and Meiselman, H.L. (2016b). An update on the roles of culture and language in designing emotion lists: English, Spanish and Portuguese. *Food Quality and Preference* 51**,** 72-76.

Vansteelandt, K., Claes, L., Muehlenkamp, J., De Cuyper, K., Lemmens, J., Probst, M., Vanderlinden, J., and Pieters, G. (2013). Variability in affective activation predicts non-suicidal self-injury in eating disorders. *European Eating Disorders Review* 21**,** 143-147.

Vastenburg, M., Romero Herrera, N., Van Bel, D., and Desmet, P. (2011). "PMRI: development of a pictorial mood reporting instrument", in: *CHI '11 Extended Abstracts on Human Factors in Computing Systems.* (Vancouver, BC, Canada: ACM).

Velasco, C., Michel, C., Woods, A.T., and Spence, C. (2016). On the importance of balance to aesthetic plating. *International Journal of Gastronomy and Food Science* 5–6**,** 10-16.

Velasco, C., Obrist, M., Petit, O., and Spence, C. (2018a). Multisensory Technology for Flavor Augmentation: A Mini Review. *Front Psychol* 9**,** 26.

Velasco, C., Obrist, M., Petit, O., and Spence, C. (2018b). Multisensory technology for flavor augmentation: A mini review. *Frontiers in Psychology* 9.

Veldhuizen, M.G., Bender, G., Constable, R.T., and Small, D.M. (2007). Trying to detect taste in a tasteless solution: Modulation of early gustatory cortex by attention to taste. *Chemical Senses* 32**,** 569-581.

Veltman, J., and Gaillard, A. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* 41**,** 656-669.

Venkatraman, V., Dimoka, A., Pavlou, P.A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H.E., Ishihara, M., and Winer, R.S. (2015). Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *Journal of Marketing Research* 52**,** 436-452.

Verastegui-Tena, L., Schulte-Holierhoek, A., Van Trijp, H., and Fiszman, B.P. (2017). Beyond expectations: The responses of the autonomic nervous system to visual food cues. *Physiology & Behavior*.

Vidal, L., Ares, G., and Jaeger, S.R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference* 49**,** 119-128.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.a.a.P. (2005). Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. *Food Quality and Preference* 16**,** 691-703.

Wackerbauer, R., Witt, A., Atmanspacher, H., Kurths, J., and Scheingraber, H. (1994). A comparative classification of complexity measures. *Chaos, Solitons & Fractals* 4**,** 133-173.

Wadhera, D., and Capaldi-Phillips, E.D. (2014). A review of visual cues associated with food on food acceptance and consumption. *Eating Behaviors* 15**,** 132-143.

Walla, P., Richter, M., Färber, S., Leodolter, U., and Bauer, H. (2010). Food-evoked changes in humans. *Journal of Psychophysiology* 24**,** 25-32.

Walla, P., Rosser, L., Scharfenberger, J., Duregger, C., and Bosshard, S. (2013). Emotion ownership: Different effects on explicit ratings and implicit responses. *Psychology* 4**,** 213-216.

Wallbott, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology* 28**,** 879-896.

Walsh, A.M., Duncan, S.E., Bell, M.A., O'keefe, S., and Gallagher, D.L. (2017). Integrating implicit and explicit emotional assessment of food quality and safety concerns. *Food Quality and Preference* 56**,** 212-224.

Walsh, E.M., and Kiviniemi, M.T. (2014). Changing how I feel about the food: experimentally manipulated affective associations with fruits change fruit choice behaviors. *Journal of behavioral medicine* 37**,** 322-331.

Wang, Q.J., Wang, S., and Spence, C. (2016a). "Turn Up the Taste": Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chemical Senses* 41**,** 345-356.

Wang, W., Den Brinker, A.C., Stuijk, S., and De Haan, G. (2016b). Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering* 64**,** 1479-1491.

Wang, X., Kumar, D., Thome, M., Cord, M., and Precioso, F. (2015). "Recipe recognition with large multimodal food dataset", in: *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).*).

Wansink, B., and Van Ittersum, K. (2013). Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *Journal of Experimental Psychology: Applied* 19**,** 320-332.

Watson, D., Clark, L.A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* 54**,** 1063-1070.

Watson, D.G., Blagrove, E., Evans, C., and Moore, L. (2012). Negative triangles: Simple geometric shapes convey emotional valence. *Emotion* 12**,** 18-22.

Wedin, L. (1972). A multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology* 13**,** 241-257.

Wei, L., Tian, Y., Wang, Y., Ebrahimi, T., and Huang, T. (Year). "Automatic webcam-based human heart rate measurements using laplacian eigenmap", in: *Asian Conference on Computer Vision*: Springer), 281-292.

Weiss, B., O'mahony, M., and Wichchukit, S. (2010). Various paired preference tests: Experimenter effect on take home choice. *Journal of Sensory Studies* 25**,** 778-790.

Weymar, M., Löw, A., Öhman, A., and Hamm, A.O. (2011). The face is more than its parts – Brain dynamics of enhanced spatial attention to schematic threat. *Neuroimage* 58**,** 946-954.

Wichchukit, S., and O'mahony, M. (2010a). Paired preference tests: 'Liking', 'Buying' and 'Take Away' preferences. *Food Quality and Preference* 21**,** 925-929.

Wichchukit, S., and O'mahony, M. (2010b). Paired preference tests:'Liking','Buying'and 'Take Away'preferences. *Food Quality and Preference* 21**,** 925-929.

Wichchukit, S., and O'mahony, M. (2011). 'Liking', 'Buying', 'Choosing' and 'Take Away' preference tests for varying degrees of hedonic disparity. *Food Quality and Preference* 22**,** 60-65.

Wierzbicka, A. (1986). Human emotions: universal or culture-specific? *American anthropologist* 88**,** 584-594.

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals.* Cambridge University Press.

Willner, P., and Healy, S. (1994). Decreased hedonic responsiveness during a brief depressive mood swing. *Journal of Affective Disorders* 32**,** 13-20.

Wilson, T.D., Lisle, D.J., Schooler, J.W., Hodges, S.D., Klaaren, K.J., and Lafleur, S.J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin* 19**,** 331-339.

Windhager, S., Slice, D., Schaefer, K., Oberzaucher, E., Thorstensen, T., and Grammer, K. (2008). Face to face: The perception of automotive designs. *Human Nature* 19**,** 331-346.

Winkielman, P., Berridge, K., and Sher, S. (2011a). Emotion, consciousness, and social behavior. *Handbook of social neuroscience***,** 195-211.

Winkielman, P., Berridge, K.C., and Sher, S. (2011b). "Emotion, consciousness, and social behavior," in *The Oxford Handbook of Social Neuroscience,* eds. J. Decety & J.T. Cacioppo. (New York: Oxford University Press), 195-211.

Winston, J.S., Gottfried, J.A., Kilner, J.M., and Dolan, R.J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *The Journal of Neuroscience* 25**,** 8903-8907.

Winton, W.M., Putnam, L.E., and Krauss, R.M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology* 20**,** 195-216.

Woods, A., Poliakoff, E., Lloyd, D., Kuenzel, J., Hodson, R., Gonda, H., Batchelor, J., Dijksterhuis, G., and Thomas, A. (2011). Effect of background noise on food perception. *Food Quality and Preference* 22**,** 42-47.

Woods, A.T., Velasco, C., Levitan, C.A., Wan, X., and Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ* 3**,** e1058.

Woodward, H.E., Treat, T.A., Cameron, C.D., and Yegorova, V. (2017). Valence and arousal-based affective evaluations of foods. *Eating Behaviors* 24**,** 26-33.

World Medical Association (2013). World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association* 310**,** 2191-2194.

Worldmedicalassociation (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists* 81**,** 14.

Wu, D., Wu, T.-I., Singh, H., Padilla, S., Atkinson, D., Bianchi-Berthouze, N., Chantler, M., and Baurley, S. (2011). "The affective experience of handling digital fabrics: Tactile and visual cross-modal effects", in: *Affective Computing and Intelligent Interaction.* (eds.) S. D'mello, A. Graesser, B. Schuller & J.-C. Martin. Springer Berlin Heidelberg).

Xu, C., Chen, L., and Chen, G. (2008). "A color bar based affective annotation method for media player", in: *Frontiers of WWW Research and Development - APWeb 2006.* (eds.) X. Zhou, J. Li, H.T. Shen, M. Kitsuregawa & Y. Zhang. (Berlin / Heidelberg: Springer).

Yao, E., Lim, J., Tamaki , K., Ishii, R., Kim, K.-O., and O'mahony, M. (2003). Structured and unstructured 9-point hedonic scales: A cross cultural study with American, Japanese and Korean consumers. *Journal of Sensory Studies* 18**,** 115-139.

Yeh, L.L., Kim, K.O., Chompreeda, P., Rimkeeree, H., Yau, N.J.N., and Lundahl, D.S. (1998). Comparison in use of the 9-Point hedonic scale between Americans, Chinese, Koreans, and Thai. *Food Quality and Preference* 9**,** 413-419.

Yoshikawa, T., Tanaka, M., Ishii, A., and Watanabe, Y. (2014). Suppressive responses by visual food cues in postprandial activities of insular cortex as revealed by magnetoencephalography. *Brain research* 1568**,** 31-41.

Yu, C.P., Samaras, D., and Zelinsky, G.J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision* 14**,** 1-16.

Yusoff, Y.M., Ruthven, I., and Landoni, M. (2013). "Measuring emotion: A new evaluation tool for very young children", in: *4th Int. Conf. on Computing and Informatics (ICOCI 2013).* Universiti Utara Malaysia).

Zampollo, F., Wansink, B., Kniffin, K.M., Shimizu, M., and Omori, A. (2012). Looks good enough to eat: How food plating preferences differ across cultures and continents. *Cross-Cultural Research* 46**,** 31-49.

Zandstra, E.H., De Graaf, C., Mela, D.J., and Van Staveren, W.A. (2000). Short- and long-term effects of changes in pleasantness on food intake. *Appetite* 34**,** 253-260.

Zandstra, E.H., De Graaf, C., Van Trijp, H.C.M., and Van Staveren, W.A. (1999). Laboratory hedonic ratings as predictors of consumption. *Food Quality and Preference* 10**,** 411-418.

Zandstra, E.H., and El-Deredy, W. (2011). Effects of energy conditioning on food preferences and choice. *Appetite* 57**,** 45-49.

Zech, H.G., Rotteveel, M., Van Dijk, W.W., and Van Dillen, L.F. (2020). A mobile approach-avoidance task. *Behavior Research Methods***,** 1-13.

Zellner, D.A., Lankford, M., Ambrose, L., and Locher, P. (2010). Art on the plate: Effect of balance and color on attractiveness of, willingness to try and liking for food. *Food Quality and Preference* 21**,** 575-578.

Zellner, D.A., Siemers, E., Teran, V., Conroy, R., Lankford, M., Agrafiotis, A., Ambrose, L., and Locher, P. (2011). Neatness counts. How plating affects liking for the taste of food. *Appetite* 57**,** 642-648.