Comparing a product-specific versus a general emoji list to measure consumers' emotional associations with chocolate and predict food choice

Joachim J. Schouteten[1], Jan Verwaeren[2], Listia Rini[1], Valérie L. Almli[3]

[1]Department of Agricultural Economics, Ghent University, Coupure Links 653, 9000 Gent, Belgium

[2]Department of Data Analysis and Mathematical Modeling, Ghent University, Coupure Links 653, 9000 Gent, Belgium

[3]Nofima, Norwegian Institute of Food, Fisheries and Aquaculture Research, N-1430 Ås, Norway

Corresponding author: Joachim J. Schouteten – Joachim.Schouteten@UGent.be

ABSTRACT

Emoji have been proposed as a way to get additional insights in how consumers perceive food products. Recent works have indicated that emoji are able to provide distinctive emotional associations with food products, regardless of whether one is using the check-all-that-apply (CATA) or the rate-all-that-apply (RATA) scaling approach. Typically, in examining emotional associations one can work with either a general list which can be used with all food products or a product-specific emotion list. To date, a comparison between the performance of a general and product-specific emoji list with adults is lacking. Moreover, it is unclear to which extent emotional data of emoji help to better predict the actual food choice of adult consumers.

Using five samples of chocolates, this study compared the use of a general list of 39 emoji with a product-specific list of 20 emoji (based upon input of 32 consumers). In total, 138 consumers assessed the samples using the general list while 136 consumers evaluated the samples with the product-specific emoji list. The RATA approach was used for the evaluation of the samples and the actual food choice was registered as participants received a snack portion of the chosen sample to take home.

Results indicated that, considering the frequency of selection, 10 emoji discriminated between the samples for both the general and product-specific lists. Similar results were obtained when considering the rating intensities. Including emoji did not lead to a significant increase in the food choice prediction regardless the type of list used. However, emoji data obtained from the product-specific emoji list was able to predict the food choice as accurate as the liking data when using the RATA intensity scores.

This study suggests that both general and product-specific emoji lists are able to generate distinguishing emotional profiles for chocolate samples. While further research is necessary with other food products and measurement methods (e.g. CATA), this study proposes that emoji measurements might be an alternative to liking data in order to be better understand of consumers' food choice.

KEYWORDS

Emotion, food, consumer, choice, rate-all-that-apply (RATA), emoji

1. Introduction

The high market failure rate of new food products indicates that merely acceptance measurements are insufficient to predict actual consumer behavior and food choice. This has led to a growing interest in going broader than acceptance by including other responses such as perceptions about the sensory attributes (Valentin, Chollet, Lelievre, & Abdi, 2012; Varela & Ares, 2012), emotional associations (Desmet & Schifferstein, 2008; King & Meiselman, 2010) and conceptualizations (Ng, Chaya, & Hort, 2013b; Thomson & Crocker, 2014).

Research about emotional associations has gained a lot of interest in the last decade. Several studies indicated that emotional measurements are able to discriminate between food products, even when the overall liking of the products is similar (King, Meiselman, & Carr, 2013; Ng, Chaya, & Hort, 2013a; Schouteten et al., 2015; Spinelli, Masi, Zoboli, Prescott, & Monteleone, 2015). Furthermore, Dalenberg et al. (2014) found that the inclusion of emotional associations helps to better predict the actual food choice of consumers. Hence, the inclusion of food-evoked emotional associations can help to improve the understanding of food choice and consumption behavior.

Self-reported measurements are mainly applied to measure consumers' emotional associations with food products (Lagast, Gellynck, Schouteten, De Herdt, & De Steur, 2017). Several self-reporting methods have been developed over the course of the years to measure emotional associations. During the task, consumers are typically asked what they feel when consuming a product or thinking about the consumption of the product (Jaeger, Cardello, & Schutz, 2013). Participants then need to select or rate all the applicable emotional associations which are mentioned in a list. The emotional list can be general (so applicable and used for all food products) e.g. the EsSense Profile™ (King & Meiselman, 2010) or specific for a product based upon prior testing with consumers, known as a product-specific list (Ng et al., 2013a). Product-specific emotional lists using emotional words have been applied to a wide range of food products such as fruit salads (Manzocco, Rumignani, & Lagazio, 2013), chocolate (Schouteten, Gellynck, et al., 2017; Thomson, Crocker, & Marketo, 2010), beer (Chaya et al., 2015), orange juice (Thomson & Crocker, 2014), coffee (Bhumiratana, Adhikari, & Chambers Iv, 2014) and yogurt (Schouteten, De Steur, Sas, De Bourdeaudhuij, & Gellynck, 2017). Other approaches to measure self-reported emotional associations include the use of terms in a wheel-format, e.g. EmoSensory® Wheel (Schouteten et al., 2015; Schouteten, Gellynck, et al., 2017), working with sentences using a semiotic approach, e.g. EmoSemio (Pierguidi, Spinelli, Dinnella, Prescott, & Monteleone, 2019; Spinelli, Masi, Dinnella, Zoboli, & Monteleone, 2014) and using animated cartoon characters, e.g. PrEmo® (Desmet, 2003; Gutjar, de Graaf, et al., 2015).

Despite the popularity of word-based approaches for examining emotional associations with tasted food products, concerns have been raised about the ecological validity as consumers seldom use words to express their food-related emotions (Köster & Mojet, 2015). Jaeger et al. (2013) indicated that several consumers found it a rather weird task to perform and that participants can even struggle to connect food stimuli to emotional words listed in a questionnaire. Therefore, there is an emerging interest in examining the potential use of emoji instead of words in food research. Vidal, Ares, and Jaeger (2016) examined the use of emoji in food-related tweets. Further research showed that emoji were helpful to examine the emotional associations linked with food products in USA and China (Ares & Jaeger, 2017; Jaeger, Vidal, Kam, & Ares, 2017). Recent research confirmed that emoji can generate discriminating emotional profiles when consumers are tasting food products (Jaeger, Lee, et al., 2017), regardless if they are using a check-all-that-apply (CATA) or rate-all-that-apply (RATA) response format (Jaeger, Lee, et al., 2018). Given that emoji are nowadays used in lots of communication methods, their use also gathered interest for food research with children. Gallo, Swaney-Stueve, and Chambers (2017b) examined children's (8-11 years old) meaning of emoji using a focus group approach. In another study, the same researchers found that emoji could be used with children to generate discriminative emotional profiles (Gallo, Swaney-Stueve, & Chambers, 2017a). The potential of obtaining discriminating profiles when children tasted food samples was confirmed in Belgium (Schouteten, Verwaeren, Lagast, Gellynck, & De Steur, 2018). In another study, the performance of a product-specific versus a general emoji list was compared using biscuit samples with a children population (Schouteten, Verwaeren, Gellynck, & Almli, 2019). That study concluded that the product-specific emoji list was better able to discriminate between the samples.

So it is established that emotional measurements contain information that is relevant to discriminate between food products and even contain information that is partly complementary to the overall liking of a food product. As a result, information that is derived from RATA questionnaires based on emoji (either using a general or a product-specific list) could potentially be used to predict the actual food choice. Schouteten, Verwaeren, et al. (2018) found that information derived from CATA questionnaires (using emoji) can indeed be used to predict the particular kind of snack that is chosen by young children. The paper concluded that information of the CATA questionnaires can (albeit rather marginally) improve prediction accuracy for food choice. But to date, no studies examined the potential contribution of emoji on the food choice prediction with adults. Furthermore, the following questions remain: (1) does rating improve food choice prediction as compared to simple applicability information as in CATA; and (2) does a product-specific list lead to statistical models with higher predictive power as compared to statistical models based on a general list?

Although the potential use of emoji in food research is gaining momentum, there has recently been a call by Jaeger, Vidal, and Ares (2021) to further examine the potential of emoji in food-related emotion research which also has been supported by Schouteten and Meiselman (2021). Prior food and beverage research using emoji to examine emotional associations with adults worked with general emoji lists, so it is unclear to which extent product-specific emoji lists have a similar performance (e.g. discriminative ability) as has been found when working with emotional words. Product-specific emotion questionnaires contain fewer items than general questionnaires as the latter encompass many items to ensure that no important items are lacking (Jaeger et al., 2013; Schouteten, Gellynck, et al., 2017; Spinelli et al., 2015). This can be beneficial when the size of the list needs to be reduced due to time usage (e.g. during mall tests) or to practical constraints, such as data collection on smaller screen sizes (e.g. smartphones). On the other hand, general questionnaires need less total project time, cost and effort as no step for the pre-selection of items is needed. Yet a longer questionnaire might induce boredom and fatigue of the participants (Jaeger et al., 2013).

This study aims to compare the performance of a product-specific emoji list with a general emoji list using the RATA response format to assess the emotional associations of adult consumers with chocolate. Hereby lies the focus on discriminatory ability and prediction of food choice, examining for the latter to which extent rating data contains additional information compared to the frequency of selection data for the food choice prediction and compare the predictive power of statistical models.

## 2. Materials and methods

### 2.1. Experimental design

This study applied a between-subjects design to compare the emotional profiling of a food product using a product-specific and a general emoji list. One group of consumers evaluated the product samples using the general emoji list of Jaeger, Lee, et al. (2017), containing both facial (33) and non-facial (6) emoji which has been applied to several food products in prior research (Table 1). Another group of consumers evaluated the same product samples but used a product-specific emoji list of facial (15) and non-facial (5) emoji generated during a pretest (see part 2.4). In total, 32 consumers were recruited for this pretest (mean age = 22 years, 58% female) at a local university campus in Ghent (Belgium) through flyers. They were only allowed to participate if they were chocolate consumers (at least once a month), used emoji when messaging/texting and had no food allergies. All emoji were in Apple iOS v.12. The study has been approved by the ethical committee of Ghent University Hospital and was conducted in accordance with the declaration of Helsinki.

Table 1. General emoji list from Jaeger, Lee et al. (2017) (emoji names are derived from Emojipedia (2021))

| | | | |
|---|---|---|---|
| 😠 angry face | 👏 clapping hand sign | 😖 confounded face | 😕 confused face |
| 😢 crying face | 😞 disappointed face | 😑 expressionless face | 😱 face screaming in fear |
| 😘 face throwing kiss | 😓 face with cold sweat | 😛 face with stuck out tongue | 😝 face with stuck out tongue and tightly closing eyes |
| 😜 face with stuck out tongue and winking eye | 😂 face with tears of joy | 😳 flushed face | 😬 grimacing face |
| 😀 grinning face | 😁 grinning face with smiling eyes | 😭 loudly crying face | 😐 neutral face |
| 👌 OK hand sign | 😔 pensive face | 😣 persevering face | 😡 pouting face |
| ❤️ red heart | 😌 relieved face | 😴 sleeping face | ☺️ smiling face |
| 😍 smiling face with heart shaped eyes | 😊 smiling face with smiling eyes | 😄 smiling face with smiling eyes and open mouth | 😎 smiling face with sunglasses |
| 😏 smirking face | 👍 thumbs up sign | 👎 thumbs down sign | 😫 tired face |
| 😒 unamused face | 😩 weary face | 😉 winking face | |

## 2.2. Participants

Participants were recruited through flyers on a local university campus in Ghent, Belgium. They were randomly assigned to either the group using a general emoji list or the group using a product-specific emoji list. Testing took place in the sensory facilities of Ghent University and participants were seated in individual booths to assess the samples. Eligibility criteria were (i) having no allergies towards nuts, gluten, milk or soy, (ii) consuming chocolate (at least once a month) and (iii) using emoji when messaging. Socio-demographic characteristics and information of internet usage, mobile devices owned and emoji usage of participants of both groups who fully completed the questionnaire can be found in Table 2. Statistical analyses showed no significant differences between the two groups for any socio-demographic or behavioral characteristic.

Table 2. Characteristics of the consumer sample for the general emoji list (n=138) and product-specific emoji list groups (n=136)

| | General emoji list (n = 138) | Product-specific emoji list (n=136) |
|---|---|---|
| Mean age, in years (S.D.) | 21.2 (3.5) | 21.5 (3.5) |
| Gender (% females) | 64.5% | 62.5% |
| Internet usage (% respondents) | | |
| Less than every two weeks | 0 | 0 |
| Once every two weeks | 0 | 0 |
| Once a week | 0 | 0 |
| Multiple times a week | 0 | 1.5 |
| Daily | 100 | 98.5 |
| Mobile devices owned (e.g. tablet/Ipad, smartphone, laptop, smartwatch) (% respondents) | | |
| 0 | 0 | 0 |
| 1 device | 0.7 | 0.7 |
| 2 devices | 59.4 | 59.6 |
| More than 2 devices | 39.9 | 39.7 |
| Emoji usage in messaging communication (% respondents) | | |
| Infrequently | 5.1 | 6.6 |
| Sometimes | 60.1 | 64.7 |
| (Almost) every time | 34.8 | 28.7 |

## 2.3. Product samples

This study opted to work with chocolate samples. Previous research indicated that chocolate consumption can influence the mood and elicit emotions such as joy and happiness (Macht & Dettmer, 2006). Furthermore, chocolate has already been used in several studies to examine consumers' emotional associations (Gunaratne et al., 2019; Schouteten, De Pelsmaeker, et al., 2018; Spinelli, Monteleone, Ares, & Jaeger, 2019; Thomson et al., 2010). Five plain milk chocolate samples were selected for this study spanning the variety of milk chocolate samples on the Belgian market regarding sensory attributes and brands. These products were selected based upon prior research carried out at Ghent University with a wider variety of chocolate samples evaluated by a trained panel using generic descriptive analysis. All samples were from the same batch and bought at local supermarkets. Nutritional information of the selected samples can be found in supplementary material (S1).

Samples were presented in transparent plastic cups labelled with a random three-digit code. Chocolate samples were served at room temperature (21 °C ± 1 °C) and serving size allowed 2–3 bites of each sample (Jaeger, Lee, et al., 2017). Product samples were presented in monadic sequence with a design balanced for presentation order and carry-over effects (Williams' design) (MacFie, Bratchell, Greenhoff, & Vallis, 1989). Participants were instructed to rinse their mouth using water between the assessment of different samples and non-salted crackers were also available for rinsing if a participant deemed it necessary.

### 2.4. Product-specific emoji list

The product-specific emoji list was developed during a pretest following the two-step procedure described in Schouteten et al. (2015). In short, a group of 32 consumers (who did not participate in the main experiment) evaluated the applicability of a list of emoji when tasting the same chocolate samples as used in the main experiment. This candidate list was based upon prior food research with adult consumers using emoji for emotional profiling when tasting food products (Jaeger, Lee, et al., 2017; Jaeger, Roigard, & Ares, 2018). Also, participants had the opportunity to add any missing emoji and provide written feedback. In a second step, the researchers made a final selection. The selection was based upon the consumers' frequency usage of emoji (≥10%), and the potential to discriminate between product samples (i.e. not same frequency for each product) (See supplementary material S2 for the results of the pretest). This is in line with previous research selecting emotional terms or emoji for developing a product-specific list (De Pelsmaeker, Schouteten, & Gellynck, 2013; Ng et al., 2013a; Schouteten et al., 2019). The final product-specific emoji list contained 20 emotions of different valence (15 positive, 3 negative and 2 neutral), arousal (3 low, 12 mediocre and 5 high) with classification based on prior research for the facial emoji (Jaeger, Roigard, Jin, Vidal, & Ares, 2019) and researchers' opinion for the non-facial emoji. Moreover, the product-specific emoji list did not only contain facial but also non-facial emoji. An overview of the product-specific emoji list can be found in Table 3.

Table 3. Product-specific emoji list for chocolate samples (emoji names are derived from Emojipedia (2021))

| | | | |
|---|---|---|---|
| 👏 clapping hand sign | 😖 confounded face | 😕 confused face | 😛 face with stuck out tongue |
| 😜 face with stuck out tongue and winking eye | 😬 grimacing face | 😀 grinning face | 😐 neutral face |
| 👌 OK hand sign | ❤️ red heart | 😌 relieved face | ☺️ smiling face |
| 😍 smiling face with heart shaped eyes | 😊 smiling face with smiling eyes | 😄 smiling face with smiling eyes and open mouth | 😎 smiling face with sunglasses |
| 😏 smirking face | 👍 thumbs up sign | 👎 thumbs down sign | 😉 winking face |

### 2.5. Questionnaire

The questionnaire was administered using EyeQuestion software (Logic8 BV, the Netherlands) v4.9.4. Participants were asked three questions related to internet and emoji usage (Jaeger, Vidal, et al., 2017): (i) how many devices they owned (desktop computer, laptop computer, tablet/iPad and/or smartphone), (ii) how often they used the Internet in general and (iii) how frequently they used emoji when sending / posting a message (Table 1).

During the sensory evaluation, participants first mentioned their overall liking of a sample on a 9-point hedonic scale (1 = dislike extremely to 9 = like extremely). Next, participants assessed the samples using a list of emoji using the rate-all-that-apply (RATA) method (Ares et al., 2014). Emoji were listed in four columns. Participants followed a two-step procedure: they were first asked to select all applicable emoji (= CATA step). Then in a second step, participants were asked to rate the intensity of those applicable emoji on a 3-point scale (1 = weak, 2 = medium, 3 = strong), with the rating appearing below the four columns containing the emoji (= RATA step). The presentation order of the emoji was randomized between the participants and also between the different samples for each participant. After the participants evaluated all the samples, they were asked to indicate of which sample they would like to receive a snack portion (chocolate wrapped in unbranded aluminum foil) to take home so that the actual food choice could be established (Schouteten, Verwaeren, et al., 2018). Participants had the opportunity to retaste the samples if needed before they made their choice.

In the last part, participants assessed the easiness ("It was easy to answer the questions about these samples using emoji") and tediousness ("It was tedious to answer the questions about these samples using emoji") of the task on a 7-point Likert scale (1 = disagree extremely to 7 = agree extremely) (Jaeger, Lee, et al., 2018). Lastly, participants recorded information regarding their gender and age. Drop-out numbers and time usage in the test were recorded to investigate eventual benefits of one approach over the other from these perspectives.

### 2.6. Data analysis

#### 2.6.1. Overall liking

Linear mixed modelling was applied to uncover significant differences in overall liking scores across the two treatments (general vs product-specific list). Treatment, sample and their interaction were specified as fixed effects, whereas consumer was specified as a random effect. Linear mixed modelling considering samples as a fixed effect and consumers as a random effect was also used to analyze the overall liking data for each treatment separately. Tukey's test was used for post-hoc comparison of means. Internal preference mapping was carried out for the two treatments (general vs product-specific emoji list) separately using XLSTAT v 2020.4.1.1032 (Addinsoft, 2020) with Principal Component Analysis (PCA) applied to the individual hedonic rates on an array of products for the participants.

#### 2.6.2. Emotional associations

Given that participants first needed to select applicable emoji and then rated the intensity on a 3-point scale, two options for analyzing these data were considered (Ares et al., 2014). First, we only considered the selection of the applicable emoji. This enables an analysis similar to CATA-questions, where a Cochran's Q test is applied to evaluate significant differences among the samples ($p \leq 0.05$). If significant differences were found for a particular emoji, pairwise comparisons were carried out between the samples using the McNemar test. Second, we considered the intensity ratings and followed the procedure recommended by Meyners, Jaeger, and Ares (2016) for RATA questions. The average score for each emoji and each sample was calculated in a 0-3 range, considering a 0 when an emoji was deemed as not applicable by a participant. Significant differences among samples were established using analysis of variance (ANOVA), followed by post-hoc comparison of means (Tukey's HSD), where $p \leq 0.05$.

Responses to task perception questions (easiness and tediousness) were analyzed using t-tests.

### 2.6.3. Links between overall liking and emotional responses

In the study design, the appreciation of a consumer towards a number of products is queried in two distinct manners. Indeed, the overall liking and the rating of the emoji can be considered as two sources of information on how the consumer perceives a specific product. It is probable that part of the information in the data is duplicated (present in both sources) to some extent and on the other hand, both sources might contain complementary information. A correlation analysis can be performed to explore this further, however, we observe that (1) a large number of emoji is used in this study and an emotional status can be represented by multiple emoji, suggesting some degree of collinearity in the responses; and (2) even though the number of respondents is above 100 and can be considered fairly large for this type of research, the statistical power of a comparison between liking and each individual emoji will be too low to be reliable. Therefore, we performed a dimensionality reduction step first that allows to reduce the dimensionality of 39, resp. 20 to a few dimensions, an approach that is similar to the one applied in prior research (Dalenberg et al., 2014; Gutjar, Dalenberg, et al., 2015; Schouteten, Verwaeren, et al., 2018). Principal Component Analysis (PCA) was used to reduce the dimensionality of data on the intensity scores, two analyses were performed for the general and product-specific lists independently. Formally, intensity ratings are organized in a data matrix such that each respondent-product combination is represented by one row and each column represents one emoji (the number of rows in this data matrix equals the product of the number of respondents and the number of products) and data are centered at the participant level. PCA is applied on that matrix. When the intensity scores are reduced to a binary format (only distinguishing to selection versus no-selection), PCA is not an optimal dimensionality reduction technique. As an alternative, Multiple Correspondence Analysis (MCA) was applied on the data. Formally, intensity data matrix is transformed to an indicator matrix (same dimensions as the original intensity matrix) and MCA is applied on that matrix as described in Pagès (2014).

When considering the intensity scores, PCA is performed (independently for the general and product-specific lists) and the scores for the two first principal components are retained as covariates. When only the frequency of selection is considered, MCA is used and the first two principal axes are retained. The principal components/axes are computed on the full dataset and scores were computed for each consumer/product combination. Biplots/factor graphs were produced to derive an interpretation for each component.

A Pearson correlation analysis was performed on both compressed versions of the dataset.

### 2.6.4. Food choice prediction

With the goal to predict food choice, we consider four settings that differ in the type of data that are used: CATA versus RATA and a general versus a product-specific list, resulting in four combinations. Note that for CATA data emoji with a rating of zero are considered irrelevant and emoji with non-zero ratings are considered relevant. For each of these settings, we build multinomial logit models in which we model the probability that a respondent will choose a particular product, using covariates based on the liking scores and the ratings (for RATA) or checks (for CATA) for each product-emoji combination. The goal of this modeling is to study how much food choice prediction can be improved by including emoji data. To prevent over-parameterization of these models due to the large number of emoji in the questionnaires, both compressed versions of the emoji data were used as mentioned in 2.6.3.

When PCA or MCA are used to reduce the dimensionality, the full multinomial regression model becomes:

$$\ln\left(\frac{P(Choice_i=j)}{P(Choice_i=K)}\right) = \alpha_j + \beta\,\mathrm{Lik_{i,j}} + \gamma_1\,\mathrm{PC}^1_{i,j} + \gamma_2\,\mathrm{PC}^2_{i,j} \text{ (Eq. 2)}$$

where $\mathrm{PC}^1_{i,j}$ is the score for the first principal component (in case of RATA data) or the first principal axis (in case of CATA data) obtained for the $i^{th}$ user and the $j^{th}$ product. $\mathrm{PC}^2_{i,j}$ represents the second principal component or axis. The multinomial logit models are fit using maximum likelihood estimation (using the R-package *mlogit* (Croissant, 2019)) and hypothesis testing was based on likelihood ratio testing.

We use a predictive modeling approach to compare both the influence of using RATA instead of CATA data and the use of a product-specific list instead of a general list. More precisely, we compute the predictive accuracy, using leave-one-out cross-validation (LOOCV), for the models in each of the four settings. To guarantee complete independence of training and test data, principal components or principal axes are recomputed in each LOOCV. Moreover, to assess the added value of PC1 and PC2 on the predictive performance, the results obtained using the full model will be compared with results obtained with reduced models in which the $\gamma$ coefficients are equal to zero. Accuracies were compared using the chi-square test for equality of proportions (available in the R *stats* package (R Core Team, 2021)).

**A note on consumer segments.** The modeling approach described above models the preference of all consumers in the same manner. However, when two or more consumer segments are present, a uniform treatment of all consumers could lead to suboptimal results. To identify potential consumer segments, hierarchical agglomerative clustering was applied on the Liking data to cluster respondents into groups with similar liking profiles. Visual inspection of the clustering tree indicated a mild clustering for both datasets, suggesting the presence of a mild degree of segmentation. To test the influence of this segmentation on the preference model, Eq. (2) was extended to incorporate the effect of clustering. However, from the obtained results we could conclude that the segmentation had no influence on the choice model. For more details on this analysis, we refer to supplementary material (S3).

# 3. Results

## 3.1. Overall liking

Table 4 presents the mean overall liking scores for the five samples, separately for both treatments. Linear mixed modelling reveals that while differences between sample means were established, there was no evidence that the type of list affected the hedonic mean scores. The interaction effect for experimental treatment x sample was non-significant ($p = 0.942$). This also shows that the two subgroups of consumers involved in the between-subjects design are on average comparable in terms of liking of the chocolate product samples. Internal preference mapping based upon the liking data for each separate list showed results in line with Table 4, with few likers of S1 and S2 and many likers of samples S3, S4 and especially S5. For the interested reader the maps can be found in supplementary material (S4).

Table 4. Mean overall liking (9-point, 1 = dislike extremely ⇔9 = like extremely) of the samples from the general list (n = 138) and product-specific list (n = 136) experiment. Different superscripts indicate significant differences in mean overall liking scores within a treatment (general / product-specific list).

| Sample | General list | Product-specific list |
|---|---|---|
| S1 | $5.2^c$ | $5.1^c$ |
| S2 | $5.1^c$ | $5.1^c$ |
| S3 | $6.0^b$ | $6.0^b$ |
| S4 | $5.9^b$ | $5.9^b$ |
| S5 | $6.6^a$ | $6.7^a$ |

$^{abc}$ different superscripts indicate significant differences between the samples of the general emoji list / product-specific emoji list based upon mixed ANOVA (samples as fixed effect & consumer as random effect). Tukey's test was used for post-hoc comparison of means (p≤0.05).

## 3.2. Emotional associations

### 3.2.1. Frequency of emoji selection

For the general emoji list, participants used on average 2.0 emoji (5.1%) to describe how they felt after consuming a chocolate sample. When working with a product-specific emoji list, 1.7 (8.5%) emoji were used on average for a chocolate sample. On aggregate level, considering all chocolate samples, usage frequencies varied for a specific emoji from 0.3% (😢) to 20.9% (😊) for the general emoji list while the usage frequencies varied between 1.6% (😜) and 20.1% (😐) for the product-specific emoji list. When comparing the usage frequencies across the samples, it appears that the number of emoji used did not differ a lot between the samples of the general list considering that the usage frequencies varied between 1.8 emoji for S4 and 2.0 emoji for S5. For the product-specific emoji list, participants used on average least emoji for the least liked sample S2 (1.5 emoji) and most for the most liked sample S5 (1.9 emoji).

For both the general and product-specific emoji lists, 10 emoji were able to discriminate between the samples (Table 5). Among these, seven emoji (😊, 😀, 😕, 😖, 👍, 👌 & 👎) are common for both lists, while six additional (general list: 😍, 😑, 😒; product-specific emoji list: 😌, ☺️, 😐) emoji were only able to discriminate between samples in one of the lists. It should be noted that two discriminating emoji of the general list (😑 , 😒) were not included in the product-specific emoji list.

Table 5. Emoji usage frequencies (in %) for the different samples of the general (n=138) and product-specific (n=136) emoji lists. Only values of emoji with significant different usage frequencies between the samples are included in this table.

| | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|
| General list (discriminative emoji out of 39 original alternatives) | | | | | |
| 😊 | $17^{b}$ | $16^{b}$ | $29^{a}$ | $22^{ab}$ | $20^{ab}$ |
| 😀 | $1^{b}$ | $7^{b}$ | $14^{a}$ | $9^{ab}$ | $14^{a}$ |
| 😕 | $13^{a}$ | $16^{a}$ | $10^{ab}$ | $12^{ab}$ | $5^{b}$ |
| 😖 | $7^{ab}$ | $11^{b}$ | $5^{ab}$ | $7^{ab}$ | $3^{a}$ |
| 👍 | $8^{c}$ | $14^{bc}$ | $14^{bc}$ | $17^{ab}$ | $23^{a}$ |
| 👌 | $4^{b}$ | $7^{ab}$ | $9^{ab}$ | $4^{b}$ | $13^{a}$ |
| 👎 | $10^{a}$ | $7^{ab}$ | $1^{c}$ | $3^{bc}$ | $1^{c}$ |
| 😍 | $6^{ab}$ | $2^{b}$ | $6^{b}$ | $6^{b}$ | $13^{a}$ |
| 😑 | $12^{a}$ | $7^{abc}$ | $2^{bc}$ | $8^{ab}$ | $2^{c}$ |
| 😒 | $8^{a}$ | $6^{a}$ | $5^{a}$ | $10^{a}$ | $0^{b}$ |
| Product-specific list (discriminative emoji out of 20 original alternatives) | | | | | |
| 😊 | $15^{c}$ | $13^{c}$ | $24^{ab}$ | $18^{bc}$ | $29^{a}$ |
| 😀 | $10^{b}$ | $10^{b}$ | $10^{b}$ | $13^{b}$ | $24^{a}$ |
| 😕 | $22^{a}$ | $26^{a}$ | $13^{b}$ | $11^{bc}$ | $5^{c}$ |
| 😖 | $12^{a}$ | $11^{a}$ | $9^{a}$ | $5^{ab}$ | $2^{b}$ |
| 👍 | $15^{bc}$ | $7^{c}$ | $21^{ab}$ | $17^{ab}$ | $25^{a}$ |
| 👌 | $7^{b}$ | $6^{b}$ | $7^{b}$ | $12^{ab}$ | $15^{a}$ |
| 👎 | $13^{a}$ | $10^{a}$ | $3^{b}$ | $7^{ab}$ | $2^{b}$ |
| 😌 | $6^{c}$ | $8^{bc}$ | $20^{a}$ | $17^{a}$ | $14^{ab}$ |
| ☺️ | $5^{b}$ | $6^{b}$ | $11^{ab}$ | $15^{a}$ | $13^{a}$ |
| 😐 | $26^{a}$ | $26^{a}$ | $21^{a}$ | $20^{a}$ | $8^{b}$ |

[abc] different superscripts indicate significant differences between the usage frequencies (in % respondents) of a certain emoji across a row based upon the McNemar test ($p \leq 0.05$)

MCA was applied to individual responses to each emoji and product configuration resulting in a MCA plot (Fig. 1a,1b) explaining 62.6% of adjusted inertia for the general emoji list and 78.4% of adjusted inertia for the product-specific emoji list. The first dimension is related to the valence of the emoji, with negatively valenced emoji situated on the left side of the plots while positively valenced emoji were located on the right side of the plots. The positioning of the emoji on the second dimension of the MCA is more difficult to interpret, but seems to be based on arousal / activation especially for the general emoji list.

Insert figure 1a,b around here

### 3.2.2. Emoji intensity scores

When considering the intensity scores, the product-specific list was slightly better in delivering discriminating emotional profiles as the product specific emoji list had 10 discriminating emoji (out of 20) while only 9 discriminating emoji (out of 39 emoji) were obtained when using a general list. It should be mentioned that 6 emoji (😀, 😕, 😖, 👍, 👌 & 👎) were significant for both lists. In addition, three more emoji were discriminating between the product samples for the general list bringing the total of discriminating emoji to 9 out of 39 original alternatives. One emoji (😍) included in both the general and product-specific emoji list was only discriminating in the general list and not in the product-specific list (Table 6).

Table 6. Average scores of the emoji for the different samples of the general (n=138) and product-specific (n=136) emoji lists. Only values of emoji with significant different average scores between the samples are included in this table.

| | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|
| General list (discriminative emoji out of 39 original alternatives) | | | | | |
| 😀 | $0.17^{ab}$ | $0.12^{b}$ | $0.33^{ab}$ | $0.23^{ab}$ | $0.36^{a}$ |
| 😕 | $0.34^{ab}$ | $0.38^{a}$ | $0.20^{ab}$ | $0.32^{ab}$ | $0.11^{b}$ |
| 😖 | $0.16^{ab}$ | $0.28^{a}$ | $0.12^{ab}$ | $0.02^{b}$ | $0.07^{b}$ |
| 👍 | $0.20^{b}$ | $0.33^{b}$ | $0.35^{ab}$ | $0.43^{ab}$ | $0.60^{a}$ |
| 👌 | $0.08^{b}$ | $0.17^{ab}$ | $0.20^{ab}$ | $0.11^{b}$ | $0.34^{a}$ |
| 👎 | $0.27^{a}$ | $0.20^{ab}$ | $0.01^{c}$ | $0.07^{bc}$ | $0.02^{c}$ |
| 😍 | $0.14^{b}$ | $0.05^{b}$ | $0.14^{b}$ | $0.15^{b}$ | $0.36^{a}$ |
| 😑 | $0.33^{a}$ | $0.18^{ab}$ | $0.04^{b}$ | $0.19^{ab}$ | $0.06^{b}$ |
| 😒 | $0.17^{ab}$ | $0.13^{ab}$ | $0.12^{ab}$ | $0.23^{a}$ | $0.00^{b}$ |
| Product-specific list (discriminative emoji out of 20 original alternatives) | | | | | |
| 😀 | $0.23^{b}$ | $0.21^{b}$ | $0.24^{b}$ | $0.30^{b}$ | $0.59^{a}$ |
| 😕 | $0.52^{ab}$ | $0.57^{a}$ | $0.29^{bc}$ | $0.24^{c}$ | $0.13^{c}$ |
| 😖 | $0.30^{a}$ | $0.25^{ab}$ | $0.17^{ab}$ | $0.14^{ab}$ | $0.04^{b}$ |
| 👍 | $0.37^{bc}$ | $0.17^{c}$ | $0.45^{ab}$ | $0.40^{abc}$ | $0.65^{a}$ |
| 👌 | $0.19^{ab}$ | $0.12^{b}$ | $0.18^{ab}$ | $0.28^{ab}$ | $0.39^{a}$ |
| 👎 | $0.32^{a}$ | $0.23^{ab}$ | $0.06^{b}$ | $0.15^{ab}$ | $0.04^{b}$ |
| 😊 | $0.33^{bc}$ | $0.27^{c}$ | $0.60^{ab}$ | $0.43^{bc}$ | $0.74^{a}$ |
| ☺️ | $0.12^{b}$ | $0.13^{b}$ | $0.27^{ab}$ | $0.38^{a}$ | $0.35^{a}$ |
| 😌 | $0.13^{b}$ | $0.19^{ab}$ | $0.42^{a}$ | $0.38^{a}$ | $0.30^{ab}$ |
| 😐 | $0.57^{a}$ | $0.60^{a}$ | $0.43^{ab}$ | $0.46^{ab}$ | $0.18^{b}$ |

$^{abc}$ different superscripts indicate significant differences between the intensity scores (ranging from 0 = not present ⇔ 3 = strong) of a certain emoji across a row. Significant differences among samples were established using analysis of variance (ANOVA), followed by post hoc comparison of means (Tukey's HSD), where $p \leq 0.05$.

PCA for emotional responses considering the intensity scores resulted in two principal components that explained 13.7% of the variance for the general emoji list and 21.1% for the product-specific emoji list (Figure 2a,2b. The first dimension of the PCA plots was associated with the valence of the emoji, with negative emoji on the left side of the plot while positively valenced emoji were situated on the right side. The second dimension of the PCA plot containing data of the general emoji list was more associated with arousal, with a positive association with emoji linked with higher emotional arousal and negative correlation with emoji associated with low emotional arousal. The second dimension of the PCA plot based upon the emoji data of the product-specific list is less clear, it seems partly related to arousal but also other factors might play a role (e.g. dominance).

Insert Figure 2a,b around here

### 3.2.3. Task perception, dropouts and time usage

No significant effect of type of list was found regarding the task perception (easiness: p = 0.643; tediousness: p = 0.592). On average, participants slightly agreed that the task was easy to perform (mean for general emoji list = 4.44 (S.D. = 1.77); mean for product-specific emoji list = 4.35 (S.D. = 1.67). Participants disagreed slightly that the task was tedious, regardless of which list was used, with an average score of 3.45 (S.D. = 1.49) for the general list and 3.55 (S.D. = 1.56) for the product-specific list. Only three dropouts were recorded for the product-specific emoji list and two for the general emoji list. Time usage was also similar for the two tests, in line with a similar number of selected emoji observed across approaches (product-specific list: 12 min. 51 s. on average (S.D. = 4 min. 19 s.), general list: 12 min. 44 s. (S.D. = 3 min. 45 s.).

### 3.3. Links between overall liking and emotional associations.

Table 7 shows the Pearson correlations between the overall liking scores and the compressed emoji data. When considering the data from the PC obtained by using PCA (frequency of selection) or MCA (intensity scores), the first PC is mainly correlated with the overall liking data.

Table 7. Pearson correlations between the compressed emoji data (PC1, PC2) and the overall liking scores. Significant ($p < 0.05$) correlations are in bold.

| | Frequency of selection | | Intensity scores | |
|---|---|---|---|---|
| | General | Product-specific | General | Product-specific |
| PC1 | **0.626** | **0.562** | **-0.442** | **0.590** |
| PC2 | **-0.139** | **-0.444** | -0.018 | **0.198** |

### 3.4. Food choice prediction

Figures 3a and 3b show the frequency distributions of the products that were finally chosen by the respondents for the experiments using the general and product-specific lists, respectively. For the general list, product S2 was only chosen 14 times, for a relative choice frequency of 10.1%, while S5 was chosen 64 times for a relative frequency of 46.4%. For the product-specific emoji list, product S2 was only chosen 11 times, for a relative choice frequency of 8.1%, while S5 was chosen 43 times for a relative frequency of 39.0%. Therefore, it can be concluded that product S5 was the most frequently chosen product in both experiments while product S2 was the least chosen. Moreover, when products are ordered according to these relative frequencies, the same ordering is obtained in both experiments.

*Insert Figure 3a,b around here*

#### 3.4.1. Prediction from frequency of emoji selection (CATA)

Figure 4 (light blue bars) shows the LOOCV accuracy that is obtained in predicting consumers' product choice using the CATA data (i.e. selection frequencies). Table 8 shows the p-values obtained when comparing the different models in a pairwise manner. As a baseline, we consider the multinomial logit model that only uses an intercept and liking scores as covariates. In the first consumer subgroup (general list group), the obtained LOOCV accuracy is 0.67 [0.59, 0.75] when only considering the liking score (CATA liking), where the LOOCV accuracy is defined as the ratio of the number of consumers for which the food choice is predicted correctly to the number of consumers in the study (Table 8). The model does not improve by adding the first principal axis of the emoji CATA data as a second covariate to the predictive model (CATA liking + PC1) as the accuracy remains at 0.66 [0.58, 0.74]. The CATA data is moderately informative for predicting food choice as the sole inclusion of the principal axis computed on the CATA data (CATA PC1) leads to an observed accuracy of 0.54 [0.46, 0.63], which is a significant drop ($p = 0.026$) compared to the predictive model only containing liking data (accuracy of 0.66 – CATA liking).

Insert Figure 4 around here

In the second consumer subgroup (product-specific list group), food choice prediction accuracy based on liking alone is 0.61 [0.53, 0.69] (Figure 4). A slight increase of the accuracy is observed when PC1 of the CATA data is added to the model from 0.61 to 0.63 [0.55, 0.71]. However, this increase is small compared to the size of the confidence intervals. In this case as well, the accuracy drops to 0.54 [0.46, 0.63] when liking is omitted as a predictor although it should be noted that the drop was not statistically significant ($p = 0.112$) (Table 9).

Table 8. Comparison of accuracies obtained for the different multinomial logit models (using LOOCV) for the general emoji list. The proportion of correctly predicted consumer preferences (identical to the accuracies shown in Figure 4) by the methods in the row/column headers are compared. The table shows p-values for testing equality of the accuracies using a z-score test for comparing proportions. p-values below the significance level are in bold. A p-value that is highlighted in bold indicates that the accuracies of the compared methods (in the row/column header) differ significantly.

| | CATA Liking | CATA Liking + PC1 | CATA Liking + PC1 + PC2 | CATA PC1 | CATA PC1 + PC2 | RATA Liking | RATA Liking + PC1 | RATA Liking + PC1 + PC2 | RATA PC1 | RATA PC1 + PC2 |
|---|---|---|---|---|---|---|---|---|---|---|
| CATA Liking | 1.000 | 1.000 | 1.000 | **0.047** | **0.026** | 1.000 | 1.000 | 0.898 | **0.005** | **0.005** |
| CATA Liking + PC1 | | 1.000 | 1.000 | 0.063 | **0.036** | 1.000 | 1.000 | 1.000 | **0.007** | **0.007** |
| CATA Liking + PC1 + PC2 | | | 1.000 | **0.047** | **0.026** | 1.000 | 1.000 | 0.898 | **0.005** | **0.005** |
| CATA PC1 | | | | 1.000 | 0.903 | **0.047** | 0.063 | 0.083 | 0.467 | 0.467 |
| CATA PC1 + PC2 | | | | | 1.000 | **0.026** | **0.036** | **0.048** | 0.628 | 0.628 |
| RATA Liking | | | | | | 1.000 | 1.000 | 0.898 | **0.005** | **0.005** |
| RATA Liking + PC1 | | | | | | | 1.000 | 1.000 | **0.007** | **0.007** |
| RATA Liking + PC1 + PC2 | | | | | | | | 1.000 | **0.010** | **0.010** |
| RATA PC1 | | | | | | | | | 1.000 | 1.000 |
| RATA PC1 + PC2 | | | | | | | | | | 1.000 |

Table 9. Comparison of accuracies obtained for the different multinomial logit models (using LOOCV) for the product-specific emoji list. The proportion of correctly predicted consumer preferences (identical to the accuracies shown in Figure 4) by the methods in the row/column headers are compared. The table shows p-values for testing equality of the accuracies using a z-score test for comparing proportions. p-values below the significance level are in bold. A p-value that is highlighted in bold indicates that the accuracies of the compared methods (in the row/column header) differ significantly.

| | CATA Liking | CATA Liking + PC1 | CATA Liking + PC1 + PC2 | CATA PC1 | CATA PC1 + PC2 | RATA Liking | RATA Liking + PC1 | RATA Liking + PC1 + PC2 | RATA PC1 | RATA PC1 + PC2 |
|---|---|---|---|---|---|---|---|---|---|---|
| CATA Liking | 1.000 | 0.803 | 0.707 | 0.326 | 0.112 | 1.000 | 0.530 | 0.616 | 1.000 | 1.000 |
| CATA Liking + PC1 | | 1.000 | 1.000 | 0.175 | **0.050** | 0.803 | 0.800 | 0.899 | 0.708 | 0.900 |
| CATA Liking + PC1 + PC2 | | | 1.000 | 0.139 | **0.037** | 0.707 | 0.899 | 1.000 | 0.617 | 0.802 |
| CATA PC1 | | | | 1.000 | 0.627 | 0.326 | 0.083 | 0.108 | 0.391 | 0.269 |
| CATA PC1 + PC2 | | | | | 1.000 | 0.112 | **0.020** | **0.027** | 0.143 | 0.087 |
| RATA Liking | | | | | | 1.000 | 0.530 | 0.616 | 1.000 | 1.000 |
| RATA Liking + PC1 | | | | | | | 1.000 | 1.000 | 0.451 | 0.614 |
| RATA Liking + PC1 + PC2 | | | | | | | | 1.000 | 0.531 | 0.706 |
| RATA PC1 | | | | | | | | | 1.000 | 0.901 |
| RATA PC1 + PC2 | | | | | | | | | | 1.000 |

### 3.4.2. Prediction from emoji intensity scores (RATA)

When considering the RATA data in the test using the general list, the observed prediction accuracies are highly similar to the accuracies obtained using the CATA data (Table 8). Differences in prediction accuracy are never significantly different when adding emoji data to the food choice prediction based upon liking data. Moreover, a drop in accuracy for the food choice prediction is observed when the liking scores are omitted when considering the RATA emoji data for the general list. For the product-specific list (listed in Table 9), the RATA data can improve the baseline accuracy 0.61 [0.53, 0.69] obtained using only liking scores to reach when adding the data of PC 1 to 0.65 [0.57, 0.73]. These results suggest that using smaller and product-specific lists is beneficial for the RATA approach. Nevertheless, it should be mentioned that the effect size is rather small and non-significant. It can also be observed that the models that only rely on RATA scores achieve accuracies that are close to the accuracies obtained when only liking is used as a covariate (0.60 [0.52, 0.69] for PC1 and 0.63 [0.54, 0.70] for PC1+PC2). This suggests that RATA data is more informative than CATA data for the product-specific list. However, here as well the accuracies do not differ significantly.

## 4. Discussion

The overall aim of this study was to compare the performance of a general emoji list with a product-specific emoji list using adults. Participants found it rather an easy task to perform and not tedious regardless of the used emoji list in line with prior research (Jaeger, Lee, et al., 2018). The number of emoji which were discriminating was similar for both emoji lists. Including the data of the emoji list in addition to the overall liking data did not improve the food choice prediction.

The overall liking score was similar for both emoji lists (interaction effect for experimental treatment x sample was non-significant ($p = 0.942$)) suggesting that the type of list does not impact a concurrent overall liking assessment. Although more research is recommended to confirm this finding, prior research with children found that the overall liking of a biscuit significantly differed between a product-specific and a general emoji list (Schouteten et al., 2019). However, it should be noted that overall the clustering of the samples according to liking was not influenced by the type of emoji list in the study of Schouteten et al. (2019), so it could be that the effect only reflected individual variations as may occur with a between-subjects design. For each sample, overall liking was assessed before the emotional assessment as suggested by King et al. (2013) to limit the potential effect of the emotional assessment on the overall liking. To date, it is still unclear if combining emoji measurements with overall liking could impact the overall liking scores but research with emotion words (King et al., 2013; Schouteten, Gellynck, et al., 2017) found little evidence that this would be the case. Nevertheless, future research should explore the potential effect of the concurrent assessment of emoji on overall liking scores and might examine if also the type of list (product-specific or general) plays a role.

Participants used slightly more emoji when working with the general list (2.0 emoji compared to 1.7) which might be influenced by the fact that the general list contained 39 emoji compared to the 20 emoji of the product-specific emoji list. The average usage frequency was higher for the product-specific emoji list (8.6%) compared to the general emoji list (5%) as the latter contained almost twice as many emoji. These numbers are low compared to prior research reporting usage frequencies between 14% and 29% when using RATA and having participants tasting food products (Ares & Jaeger, 2017; Jaeger, Lee, et al., 2018). It could be that the method of the response format used in this study, that is to say first checking an applicable emoji (CATA step) and then rating the intensity in a second step (RATA step), led to a lower usage frequency. Also the fact that only food products of a single food category were used in this study could have played a role as more diverse foods might result to larger emotion differences expressed by more emotions. Further, the use of RATA asks a higher cognitive involvement and prolongs the task, so it could be that some participants applied a satisfying answer strategy. It might be interesting to delve deeper in the impact of the response format and for instance use eye tracking to gain additional insights in how participants check & rate applicable emoji. In line with the answers provided about the easiness and tediousness of the tasks, the low frequency of use could also be caused by some participants who found it hard to assess / indicate their emotions. Prior CATA research with children in Belgium comparing a general with a product-specific emoji list also reported higher usage frequencies (15.2% for the product-specific emoji list and 7.5% for the general emoji list) (Schouteten et al., 2019).

In line with prior studies, this research found that emoji are able to discriminate between food products when using the RATA response format (Ares & Jaeger, 2017; Jaeger, Lee, et al., 2018). The present study adds that emoji are even able to discriminate between products of the same food category with similar flavour, as this study only used plain milk chocolate samples. Research of Jaeger, Lee, et al. (2018) on the other hand, used different types of foods (soft cheese, cucumber, cereal bar and cake) or products of distinctive flavour characteristics (muesli bars with apricot, apple and raspberry, apple crumble and apple flavour).

For both emoji lists, about 10 emoji were able to discriminate between the food product samples. In the case of the general list there was a slight effect of the CATA versus RATA analysis method, where 10 emoji allowed sample discrimination when only the emoji applicability was considered (CATA style) while only nine emoji were discriminative based on intensity scores (RATA approach) (😊 was no longer discriminative). Further, it is interesting to mention that the discriminating emoji were largely the same for both lists but there were still some differences as 😑, 😍, 😒 were able to discriminate only for the general list and 😐, 😌, ☺️ were only discriminating when the product-specific emoji list was used. Moreover, none of the discriminating emoji were able to discriminate between equally liked samples. This contrasts with the results of Schouteten et al. (2019) who found that emoji from the product-specific emoji list were better able to discriminate between equally liked samples compared to a general emoji list when children evaluated biscuits. It could be that equally liked products in this study had rather similar sensory properties. It has been suggested that a product's emotional profile relies on its specific sensory properties (Spinelli & Jaeger, 2019) so if the equally liked samples had similar sensory properties, this might explain the inability of the emoji to discriminate between those equally liked samples. Also, the fact that the product-specific emoji list contained predominantly positive emoji might have impaired the ability to discriminate between lesser-liked samples, although it should be noted that the general emoji list (containing a larger number of negative emoji) only had one more discriminating negative emoji than the product-specific emoji list. Future research could opt to consider the balance of positive/negative/neutral valence as a selection criterion when selecting emoji for a product-specific emoji list. Furthermore, attention should be paid that also non-facial emoji are able to discriminate between the chocolate samples when using the RATA response format, for analyzing it as frequency of use and as intensity data. Jaeger, Lee, et al. (2017) previously reported that such emoji are able to discriminate, but their research was carried out using the CATA response format. However, it is recommended that further research examine the meanings of non-facial emoji.

Prior research by Schouteten et al. (2019) examining emotional associations using emoji with biscuits consumption of children found that the 5 food samples were mainly sorted based on the emotional valence in an MCA plot. In the present study, the first CA dimension was predominantly valence-driven and accounted for about 50% (general emoji list) to 67% (product-specific list) of the variance. The second dimension was mainly based on the arousal /activation associated with the emoji and only explained about 11% of the total inertia, which was similar for both emoji lists. The higher explained variance by the product-specific emoji list is interesting to note, which might be due to fact that the product-specific emoji list contained less emoji. However, however the study of Schouteten et al. (2019) worked with 33 emoji which resulted in about 83% of the variance explained in the first dimension. The two first dimensions of the PCA plots based upon the data of the intensity scores only explained 13.7% (general emoji list) and 21.1% (product-specific emoji list) of the variance. This is rather low compared to prior research finding that the first two principal components explained over 80% of the total variance for the emotional associations based upon a RATA-response format with 41 emoji for four kefir beverages (Pinto et al., 2020). It could be that the lower usage frequency played a role, as the study of Pinto et al. (2020) reported an average use of 5.7 emoji out of 51 emoji while on average about 1.7 to 2.0 emoji were used in this study. Future research could opt to work with another response format, e.g. mandatory check of an emoji is applicable with yes/no followed by rating as this might result in higher emoji usage frequencies (Ares & Jaeger, 2017).

The multinomial logit choice models confirm that perceived liking remains the most important predictor for food choice. The information extracted from the general emoji list did not lead to a significant increase of the choice model for the chocolate. When the information of the product-specific emoji list complements the overall liking information, an increase was found especially when using the intensity scores albeit that this increase was not significant. This resulted then in a model with an accuracy of 65%. Prior research already found that including information of emoji improved the choice prediction for speculoos biscuits (accuracy of 66.4%) when working with a children's population compared to the prediction based on solely the liking scores (accuracy of 62.4%) (Schouteten, Verwaeren, et al., 2018). Research from Dalenberg et al. (2014) reported that the best model (based upon liking data and PC1 from PrEmo) had an accuracy of 54.4%. while research from Gutjar, Dalenberg, et al. (2015) established an accuracy of 44.7% for the best fitting model containing the liking and valence data (based upon EsSense Profile®).

This study worked with the RATA response format, so it is recommended that more research is carried out with other products and also with the CATA response formats to confirm these findings. Furthermore, it seems that the added value in food choice prediction is only present when working with a product-specific emoji list. It should be examined if that is due to the fact that the list is product-specific or if the lower number of emoji played a role (or a combination of both). Moreover, the fact that none of the discriminating emoji was able to discriminate between equally liked samples had potentially an impact on the relatively small effects of emoji data on choice prediction. Furthermore, it should be noted that the food choice prediction using only the intensity scores of the emoji is similar to the food choice prediction of the liking scores for the product-specific emoji list. Although further research is recommended, it shows that the emoji data might be in some cases as meaningful as liking data to predict food choice.

When interpreting the results of the food choice prediction, one should consider that the limited additional value could be due to several reasons. First of all, the predictive performance using the models that only rely on liking data was quite high compared to prior research examining the added value of emotional data using an adult population (Dalenberg et al., 2014; Gutjar, Dalenberg, et al., 2015) which makes it harder for emoji data to improve the food choice prediction. Further, the number of samples and also the fact that these samples were from a single food-product category could have played a role. Also the emoji list and the selection might have had an impact, with more research recommended to discover emoji which are able to discriminate between equally liked products. Lastly, also the response format might have play a role as prior research showed that the response format could have an impact on the emoji usage frequencies by participants (Ares & Jaeger, 2017).

This study worked with five milk chocolate samples, selected to span the market regarding sensory attributes and brands. Although this number is in line with prior research studying the potential of emotional measurements to predict food choice (Dalenberg et al., 2014; Gutjar, Dalenberg, et al., 2015; Schouteten, Verwaeren, et al., 2018), future research could examine the impact of a larger number of samples when comparing the performance of the two emoji-list approaches. Nevertheless, as emotions are expected to be elicited from the products' sensory properties and/or possibly product recognition/familiarity (Ng et al., 2013b; Schouteten, De Steur, et al., 2017; Spinelli et al., 2015; Spinelli et al., 2019), little new information would be expected from a larger number of samples since these were already selected to span the range of variations on the market. Furthermore, one need to consider the tediousness of the task and also bear the sensory satiety in mind when the experiment is carried out in a single session as that is a prerequisite for determining the food choice when products need to be tasted by the participants.

A between-subjects design was used for this study, in line with other studies examining the performance of two methods to measure consumers' emotional associations of food products (Ng et al., 2013a; Schouteten et al., 2019; Spinelli et al., 2014). Therefore, within-subject factors which may play a role when consumers assess samples under different conditions were eliminated (e.g. carry-over effects due to a first assessment). Future research might opt to apply a within-subject design although it should be mentioned that both consumer panels used in this study were highly similar regarding their socio-demographics, emoji use, and product preferences.

This study worked with a convenience sample mainly consisting of younger people. Future research might use a more representative sample of the general population. Nonetheless, it should be noted that prior research found that younger and older people don't largely differ in the adoption (Evans, 2017) nor use of emoji (Gallud, Fardoun, Andres, & Safa, 2018). Also, Jaeger, Xia, et al. (2018) found that the interpretation of emoji and their usage for describing food-related emotional associations do not seem to be largely influenced by age, gender, or the frequency of use of emoji in digital communications.

RATA data were analyzed in this study using both the frequency of use and intensity scores (Ares et al., 2014). Both analysis methods were used in order to provide a clearer picture of the data. It should be noted that future research might set up a study with on one hand data collected using a CATA format and on the other hand the RATA format for intensity scores, when the goal is to compare the performance of both response formats. One needs to bear in mind that participants in this study first needed to check all the applicable emoji and in a second step rate the intensities, which advocates the use of both analysis methods used in this study.

Dropout rates and time usage on the task were similar across approaches despite the difference in number of terms. Previous research using eye-tracking measures reported greater sustained visual attention to the task in the case of a longer CATA list, but shorter attention per term (Ares, Antúnez, Giménez, & Jaeger, 2015). The equal time usage is also in line with the observed equal frequency selections of emoji. As the selected emoji had to be rated, it is possible that participants adopted a satisficing strategy to limit their effort, thus limiting time usage also for the longer list (Ares et al., 2014; Jaeger, Lee, et al., 2018). Our data provide therefore no evidence of a time or engagement advantage/disadvantage in favor of one list or the other. It is however a fact that a shorter list will save screen space and thus better fit smaller screen formats such as smartphones, which are more and more often used for data collection in consumer studies nowadays (York & Poynter, 2018). Further, one need to consider that the product-specific emoji list is shorter than the general list which is quite common as general emoji (or emotion) lists contain many items to ensure that no important items are lacking (Jaeger et al., 2013; Schouteten, Gellynck, et al., 2017; Spinelli et al., 2015).

In conclusion, this study aimed to compare the performance of a general and a product-specific emoji list with milk chocolate samples using the RATA response format both considering the frequency of selection and the intensity scores. It showed that a product-specific emoji list is valid and slightly better in discrimination performance compared to a general list when considering the intensity scores. It also showed that non-facial emoji can be discriminating between tasted food samples of rather similar composition. Overall liking remains the best predictor of food product choices. However, emoji data considering the intensities obtained from a food-product specific emoji list was as accurate as the liking data to predict food choice. Future research may further examine the potential benefits of a product-specific list in the case of testing products within a single food category. Future research may also examine if emotional profiles help improving product choice prediction after product exposure over time, rather than in a single test, as liking may become relatively less important for choice predictions than emotions after a longer product experience.

## Acknowledgments

J.J. Schouteten acknowledges FWO (Research Foundation – Flanders, grant FWO V416118N) for enabling the cooperation with Nofima for this study. V.L. Almli acknowledges the Norwegian Fund for Research Fees for Agricultural Products (FFL) (FoodForFuture, project nr. 314318).

## Statement regarding conflicts of interest

All authors declare no conflicts of interest.

## CRediT author statement

JJS, JV, VLA: Conceptualization, Methodology, Formal analysis, Writing – Original draft, Writing – Review & Editing. LR: Investigation, Data curation

References

Ares, G., Antúnez, L., Giménez, A., & Jaeger, S. R. (2015). List length has little impact on consumers' visual attention to CATA questions. *Food Quality and Preference, 42*, 100-109.

Ares, G., Bruzzone, F., Vidal, L., Cadena, R. S., Giménez, A., Pineau, B., et al. (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., & 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.

Bhumiratana, N., Adhikari, K., & Chambers Iv, E. (2014). The development of an emotion lexicon for the coffee drinking experience. *Food Research International, 61*(0), 83-92.

Chaya, C., Eaton, C., Hewson, L., Vázquez, R. F., Fernández-Ruiz, V., Smart, K. A., et al. (2015). Developing a reduced consumer-led lexicon to measure emotional response to beer. *Food Quality and Preference, 45*, 100-112.

Croissant, Y. (2019). mlogit: Multinomial logit models [Computer software manual]. In.

Dalenberg, J. R., Gutjar, S., ter Horst, G. J., de Graaf, K., Renken, R. J., & Jager, G. (2014). Evoked Emotions Predict Food Choice. *PLoS ONE, 9*(12), e115388.

De Pelsmaeker, S., Schouteten, J., & Gellynck, X. (2013). The consumption of flavored milk among a children population. The influence of beliefs and the association of brands with emotions. *Appetite, 71*, 279-286.

Desmet, P. M. A. (2003). Measuring emotion; development and application of an instrument to measure emotional responses to products. In M. A. Blythe, A. F. Monk, K. Overbeeke & P. C. Wright, *Funology: from usability to enjoyment*. Dordrecht: Kluwer.

Desmet, P. M. A., & Schifferstein, H. N. J. (2008). Sources of positive and negative emotions in food experience. *Appetite, 50*(2-3), 290-301.

Emojipedia. (2021). In.

Evans, V. (2017). *The emoji code: How smiley faces, love hearts and thumbs up are changing the way we communicate*: Michael O'Mara Books.

Gallo, K. E., Swaney-Stueve, M., & Chambers, D. H. (2017a). Comparing visual food images versus actual food when measuring emotional response of children. *Journal of Sensory Studies, 32*(3), e12267.

Gallo, K. E., Swaney-Stueve, M., & Chambers, D. H. (2017b). A focus group approach to understanding food-related emotions with children using words and emojis. *Journal of Sensory Studies, 32*(3), e12264.

Gallud, J. A., Fardoun, H. M., Andres, F., & Safa, N. (2018). A study on how older people use emojis. In, *Proceedings of the XIX International Conference on Human Computer Interaction*.

Gunaratne, T. M., Gonzalez Viejo, C., Fuentes, S., Torrico, D. D., Gunaratne, N. M., Ashman, H., et al. (2019). Development of emotion lexicons to describe chocolate using the Check-All-That-Apply (CATA) methodology across Asian and Western groups. *Food Research International, 115*, 526-534.

Gutjar, S., Dalenberg, J. R., de Graaf, C., de Wijk, R. A., Palascha, A., Renken, R. J., et al. (2015). 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., et al. (2015). The role of emotions in food choice and liking. *Food Research International, 76*(2), 216-223.

Jaeger, S. R., Cardello, A. V., & Schutz, H. G. (2013). Emotion questionnaires: A consumer-centric perspective. *Food Quality and Preference, 30*(2), 229-241.

Jaeger, S. R., Lee, S. M., Kim, K.-O., Chheang, S. L., Jin, D., & Ares, G. (2017). Measurement of product emotions using emoji surveys: Case studies with tasted foods and beverages. *Food Quality and Preference, 62*, 46-59.

Jaeger, S. R., Lee, S. M., Kim, K.-O., Chheang, S. L., Roigard, C. M., & Ares, G. (2018). CATA and RATA questions for product-focused emotion research: Five case studies using emoji questionnaires. *Food Quality and Preference, 68*, 342-348.

Jaeger, S. R., Roigard, C. M., & Ares, G. (2018). 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., Roigard, C. M., Jin, D., Vidal, L., & Ares, G. (2019). Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research. *Food Research International, 119*, 895-907.

Jaeger, S. R., Vidal, L., & Ares, G. (2021). Should emoji replace emotion words in questionnaire-based food-related consumer research? *Food Quality and Preference*, 104121.

Jaeger, S. R., Vidal, L., Kam, K., & Ares, G. (2017). 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., & Ares, G. (2018). 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.

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

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

Köster, E. P., & Mojet, J. (2015). From mood to food and from food to mood: A psychological perspective on the measurement of food-related emotions in consumer research. *Food Research International, 76*(2), 180-191.

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

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies, 4*(2), 129-148.

Macht, M., & Dettmer, D. (2006). Everyday mood and emotions after eating a chocolate bar or an apple. *Appetite, 46*(3), 332-336.

Manzocco, L., Rumignani, A., & Lagazio, C. (2013). Emotional response to fruit salads with different visual quality. *Food Quality and Preference, 28*(1), 17-22.

Meyners, M., Jaeger, S. R., & Ares, G. (2016). On the analysis of Rate-All-That-Apply (RATA) data. *Food Quality and Preference, 49*, 1-10.

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

Ng, M., Chaya, C., & Hort, J. (2013b). The influence of sensory and packaging cues on both liking and emotional, abstract and functional conceptualisations. *Food Quality and Preference, 29*(2), 146-156.

Pagès, J. (2014). *Multiple factor analysis by example using R*: CRC Press.

Pierguidi, L., Spinelli, S., Dinnella, C., Prescott, J., & Monteleone, E. (2019). Individual differences in perceived complexity are associated with different affective responses to alcoholic cocktails. *Food Quality and Preference, 76*, 47-59.

Pinto, V. R. A., Teixeira, C. G., Lima, T. S., Prata, E. R. B. D. A., Vidigal, M. C. T. R., Martins, E., et al. (2020). Health beliefs towards kefir correlate with emotion and attitude: A study using an emoji scale in Brazil. *Food Research International, 129*, 108833.

Schouteten, J. J., De Pelsmaeker, S., Juvinal, J., Lagast, S., Dewettinck, K., & Gellynck, X. (2018). Influence of sensory attributes on consumers' emotions and hedonic liking of chocolate. *British Food Journal, 120*(7), 1489-1503.

Schouteten, J. J., De Steur, H., De Pelsmaeker, S., Lagast, S., De Bourdeaudhuij, I., & 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., De Steur, H., Sas, B., De Bourdeaudhuij, I., & Gellynck, X. (2017). The effect of the research setting on the emotional and sensory profiling under blind, expected, and informed conditions: A study on premium and private label yogurt products. *Journal of Dairy Science, 100*(1), 169-186.

Schouteten, J. J., Gellynck, X., De Bourdeaudhuij, I., Sas, B., Bredie, W. L. P., Perez-Cueto, F. J. A., et al. (2017). Comparison of response formats and concurrent hedonic measures for optimal use of the EmoSensory® Wheel. *Food Research International, 93*, 33-42.

Schouteten, J. J., & Meiselman, H. L. (2021). The potential of using emoji instead of words when measuring emotions associated with food. *Food Quality and Preference*, 104182.

Schouteten, J. J., Verwaeren, J., Gellynck, X., & Almli, V. L. (2019). Comparing a standardized to a product-specific emoji list for evaluating food products by children. *Food Quality and Preference, 72*, 86-97.

Schouteten, J. J., Verwaeren, J., Lagast, S., Gellynck, X., & De Steur, H. (2018). Emoji as a tool for measuring children's emotions when tasting food. *Food Quality and Preference, 68*, 322-331.

Spinelli, S., & Jaeger, S. R. (2019). What do we know about the sensory drivers of emotions in foods and beverages? *Current Opinion in Food Science, 27*, 82-89.

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

Spinelli, S., Masi, C., Zoboli, G. P., Prescott, J., & Monteleone, E. (2015). Emotional responses to branded and unbranded foods. *Food Quality and Preference, 42*, 1-11.

Spinelli, S., Monteleone, E., Ares, G., & Jaeger, S. R. (2019). Sensory drivers of product-elicited emotions are moderated by liking: Insights from consumer segmentation. *Food Quality and Preference, 78*, 103725.

Thomson, D. M. H., & Crocker, C. (2014). Development and evaluation of measurement tools for conceptual profiling of unbranded products. *Food Quality and Preference, 33*, 1-13.

Thomson, D. M. H., Crocker, C., & Marketo, C. G. (2010). Linking sensory characteristics to emotions: An example using dark chocolate. *Food Quality and Preference, 21*(8), 1117-1125.

Valentin, D., Chollet, S., Lelievre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science and Technology, 47*(8), 1563-1578.

Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*(2), 893-908.

Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference, 49*, 119-128.

York, S., & Poynter, R. (2018). Global mobile market research in 2017. In, *Mobile Research*: Springer.

Figures

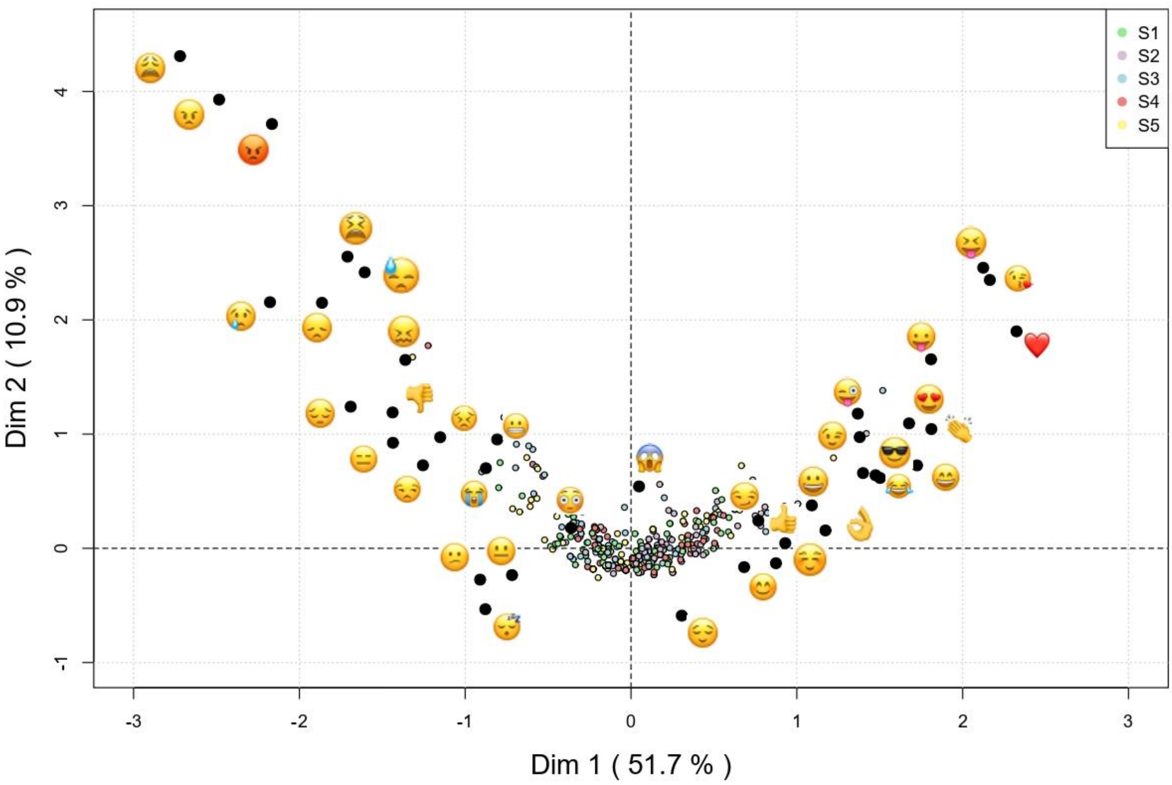


**Figure 1a. Multiple Correspondence Analysis (MCA) for the liking (supplementary variable) and emotional associations of the five samples for the general emoji list**

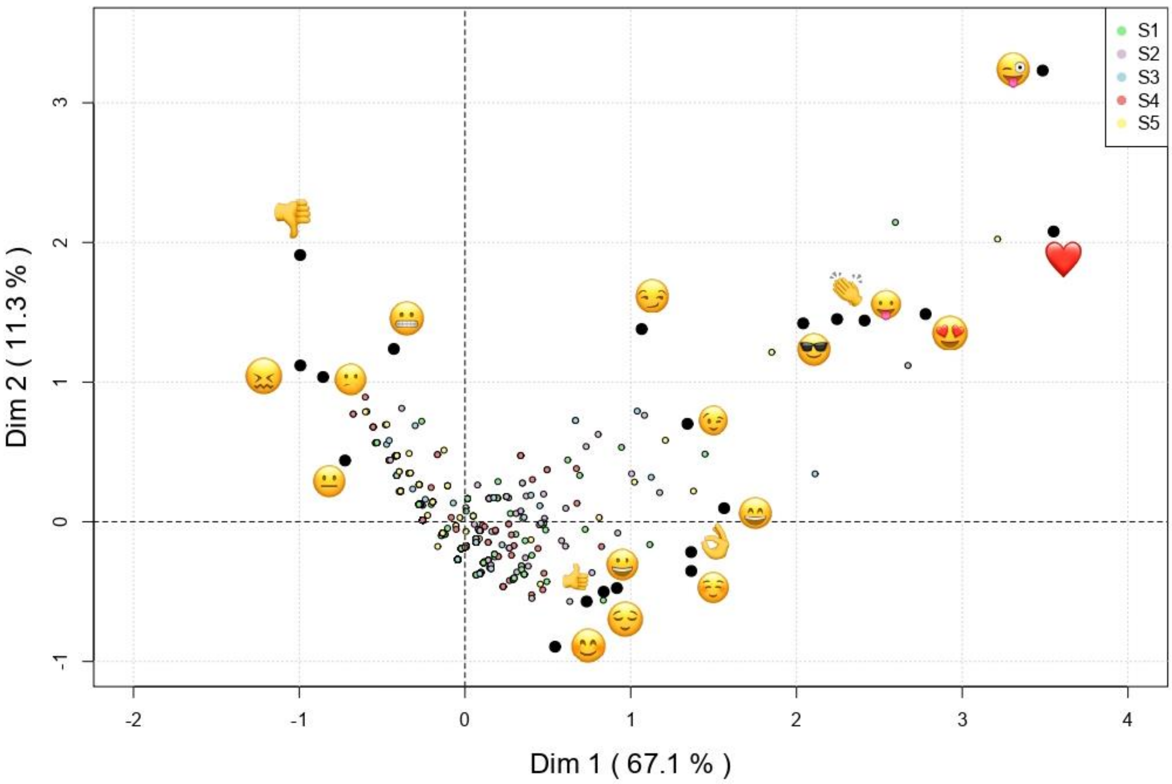


**Figure 1b. Multiple Correspondence Analysis (MCA) for the liking (supplementary variable) and emotional associations of the five samples for the product-specific emoji list**

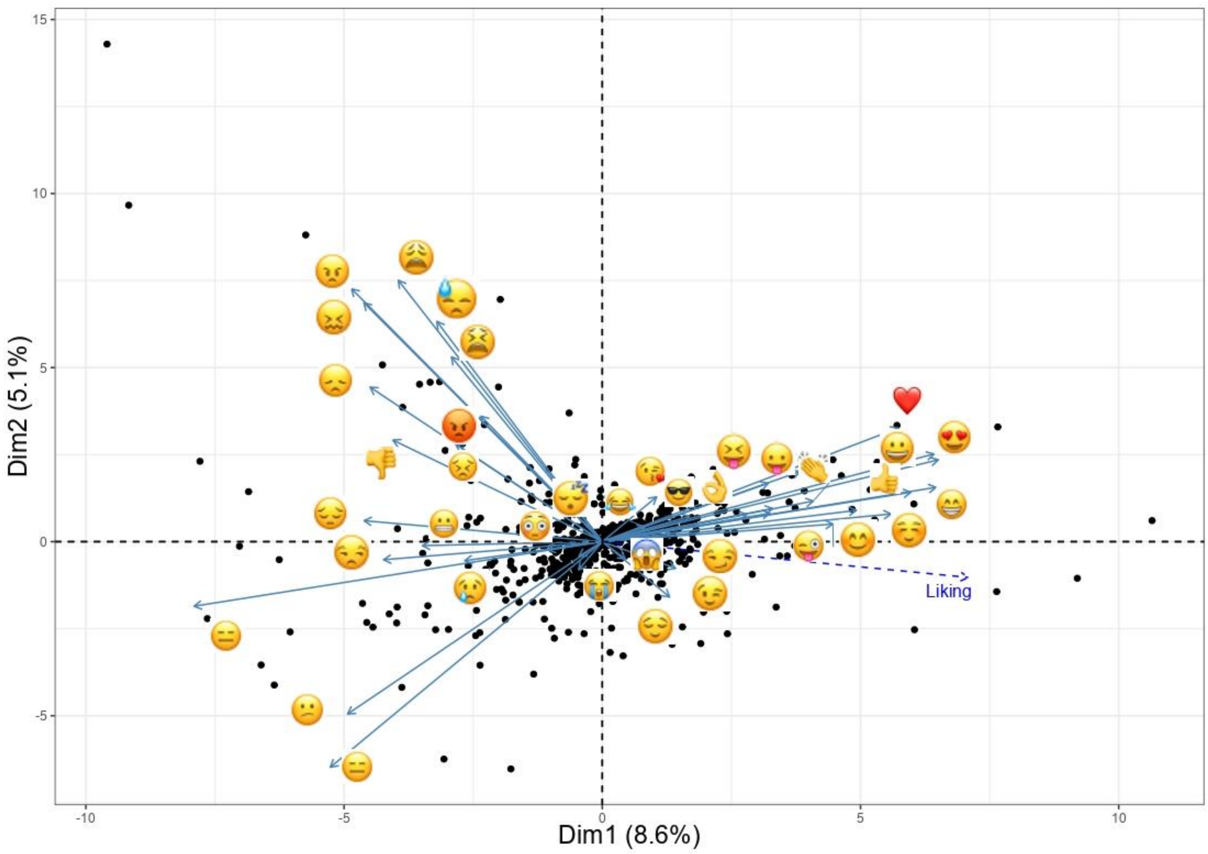


**Figure 2a. Principal Component Analysis (PCA) biplot for the liking (supplementary variable) and emotional associations of the five samples for the general emoji list**

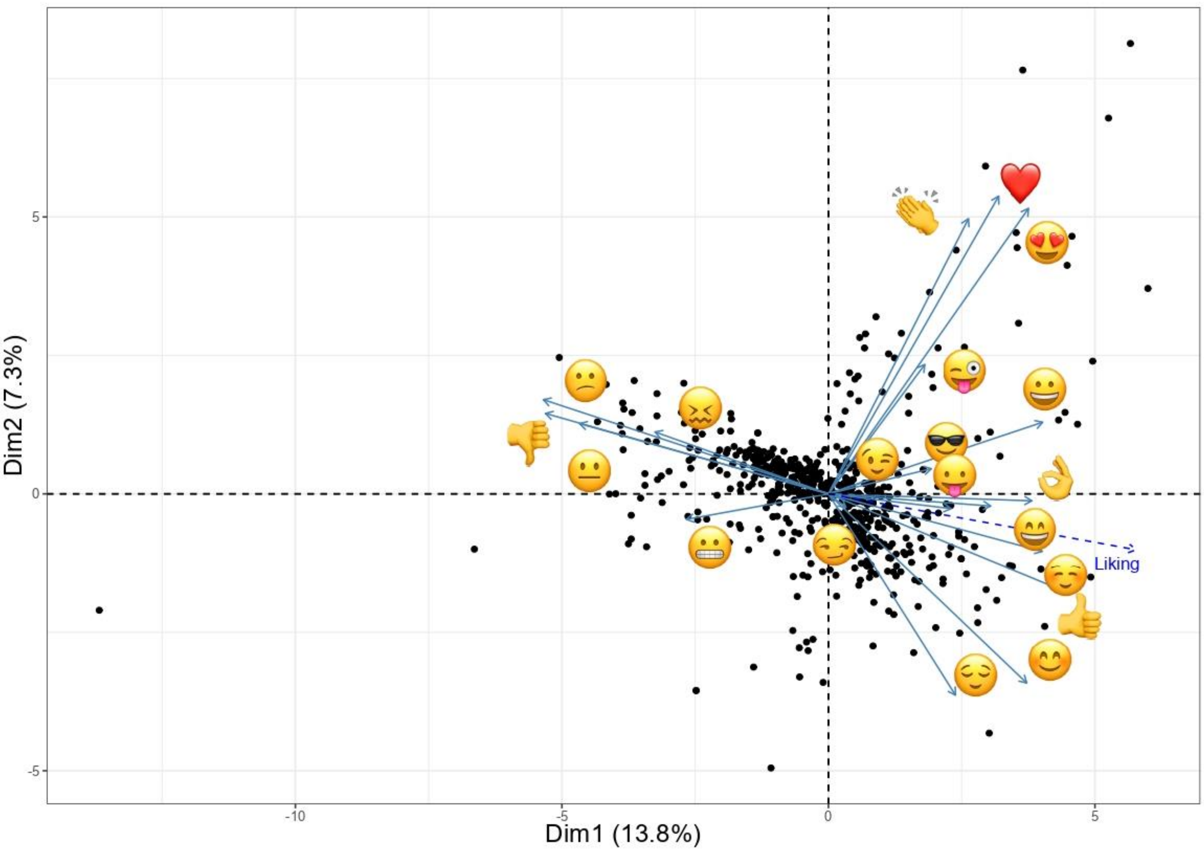


**Figure 2b. Principal Component Analysis (PCA) biplot for the liking (supplementary variable) and emotional associations of the five samples for product-specific emoji list.**

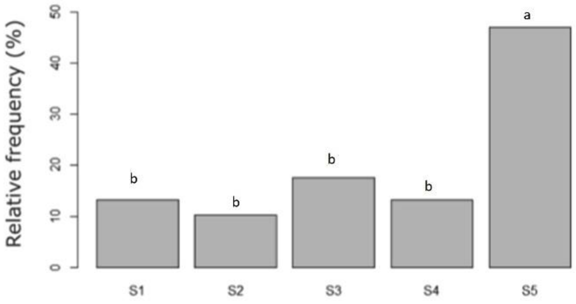


**Fig 3a. Frequency distribution of food choice for the experiment using the general list.**

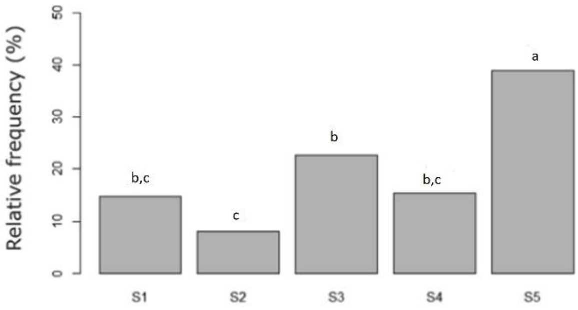


**Fig 3b : Frequency distribution of food choice for the experiment using the product-specific list.**

[abc] different superscripts indicate significant differences in the chosen food product between the samples within the same list (general / product-specific) based upon the chi² test considering p≤0.05

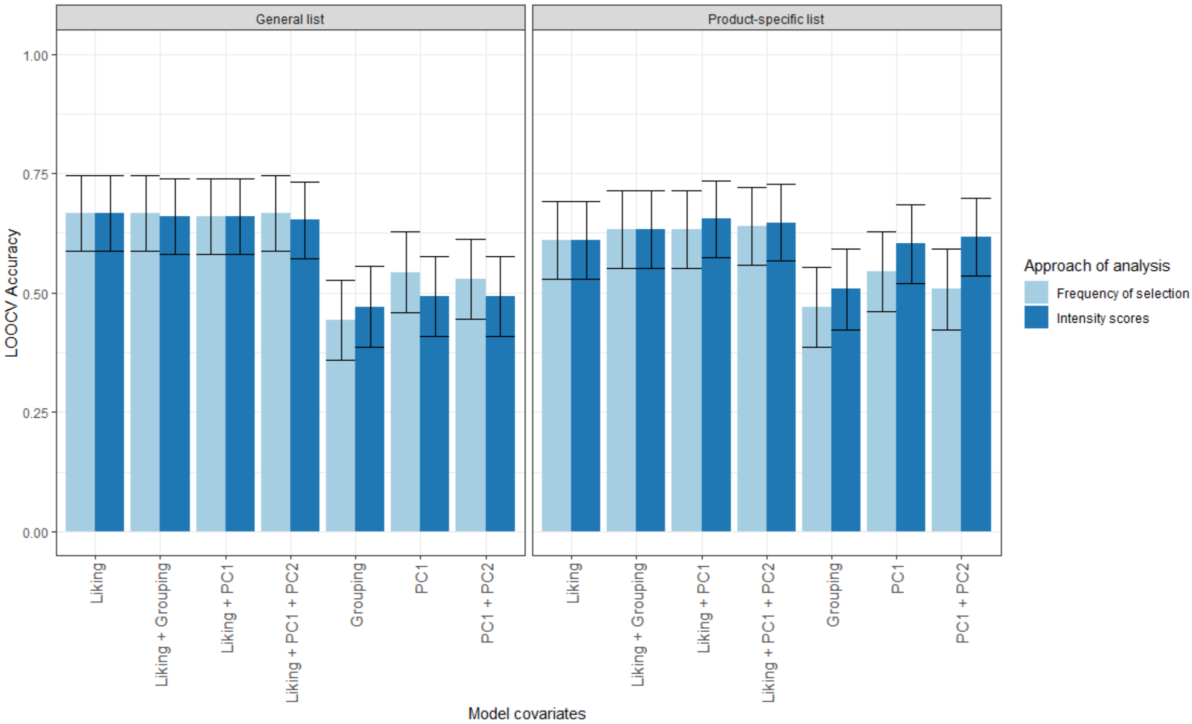


***Figure 4. Barplot showing the influence of the list that is used and the questionnaire type on the leave-one-out cross-validation (LOOCV) accuracy for five multinomial logit models. LOOCV allows for an unbiased prediction estimate, because the choice of every single individual is predicted independently from all other individuals. The error bars show 95% confidence intervals for the accuracies.***