Baseline adjustment in CATA-based emotion research: Methodological value and insights into emotional carryover

Jiahui Song[1,2,3], Yixun Xia[1,2,3], Charles Spence[4], Ziyuan Wang[5], Limin Qu[5], Juan Wang[5], & Fang Zhong[1,2,3*]

[1] School of Food Science and Technology, Jiangnan University, Wuxi, 214122, Jiangsu, China

[2] Science Center for Future Foods, Jiangnan University, Wuxi 214122, Jiangsu, China

[3] Key Laboratory of Food Sensory Science and Technology, China National Light Industry, Wuxi, 214122, China

[4] Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, New Radcliffe HouseLife & Mind Building, Oxford, OX12 6BW3PS, UK

[5] CLASSYKISS Dairy (Shenzhen) Co., Ltd

Corresponding authors:

*Fang Zhong: fzhong@jiangnan.edu.cn

Abstract

Consumers‘ baseline emotional states prior to food evaluation may influence their subsequent emotional responses, yet this is rarely considered in product research. A within-subjects experimental design was used to compare emotional responses obtained with and without baseline measurement. Check-All-That-Apply (CATA) was used to capture emotional responses to samples and to measure baseline states prior to tasting across two food product categories: yogurt (n=129) and shrimp sticks (n=134). Baseline adjustment revealed that food not only induced specific emotions but also diminished them (e.g., unsweetened yogurt reduced ‘Happy’ by 17.2%). Importantly, while baseline adjustment did not enhance sample discrimination, it uniquely helped discriminate negative emotions (e.g., ‘Disgusted’, ‘Worried’) and enhanced discriminability amongst those consumers experiencing negative emotion. This enhanced sensitivity to negative emotions is explained by their greater persistence: negative and high-arousal emotions exhibited the highest cross-sample repeatability and did not fully recover after a 2-minute rest (baseline-to-baseline agreement = 0.71–0.78). These results reflect the presence of cross-sample interference, which baseline adjustment appeared to eliminate. These findings highlight the need to incorporate baseline assessments into sequential sample testing to obtain a more accurate emotional profile.

Keywords

Emotional Response; Baseline Emotion; Check-All-That-Apply (CATA); Food; Valence; Arousal

1. Introduction

In sensory and consumer research, there has been growing interest in measuring emotional responses to food products. Such measurements provide deeper insights into the experiences associated with products (Thomson et al., 2010), helping to differentiate products with similar liking scores (Schouteten et al., 2015; Spinelli et al., 2015) and improve the predictions of consumers' purchase behavior (Dalenberg et al., 2014; Gutjar et al., 2015). One of the most popular approaches to food-related emotion research is the Check-All-That-Apply (CATA) method (Gupta et al., 2021) **(Souza-Coutinho et al., 2020) (Gunaratne et al., 2019), which asks consumers to select terms describing their feelings when tasting samples** (King & Meiselman, 2010), using a set of emotional terms such as the EsSense25 (Nestrud et al., 2016), as well as symbols like emojis (Jaeger, Lee, et al., 2018)

However, a critical yet often overlooked factor in food-evoked emotion research is the consumer's baseline emotional state prior to product evaluation. Extensive research indicates that pre-existing emotional states may influence subsequent emotional responses to stimuli. For instance, positive baseline emotional states tend to correlate with the selection of positive emotional terms, whereas negative baseline emotional states tend to correspond with the selection of negative emotional terms instead. This pattern of results has been observed across various settings, including people's reactions to wine (Danner et al., 2016), novel foods (Kaneko et al., 2020), film (Negrão et al., 2022), music (Vuoskoski & Eerola, 2011) and facial expressions (Bouhuys et al., 1995). Moreover, baseline emotional states may also influence sample discrimination based on emotion. Negrão et al. (2022) found that those individuals experiencing negative emotions tend to exhibit smaller emotional changes after viewing a film. However, despite evidence from other fields, it remains unclear whether and how consumers' baseline emotional states influence their emotional responses measured by CATA questionnaires.

A further arises from the persistence of emotions (Adolphs & Andler, 2018), a fundamental characteristic of affective responding. Research has demonstrated that emotions do not terminate immediately upon stimulus cessation but can persist for a considerable duration, as evidenced in studies using video and music (Gomez et al., 2009; Kuijsters et al., 2016; Ribeiro et al., 2019). In sensory and consumer tests, however, multiple samples are typically evaluated sequentially within a single session. Under such conditions, the emotional state elicited by a preceding sample may still be present when the next sample is evaluated, thereby potentially carrying over and biasing responses to subsequent samples. Although balanced presentation orders such as Williams Latin square designs are commonly usemployed to mitigate such interference, these designs do not eliminate the confounding influence of carryover effect (Lawless & Heymann, 2010). How such emotional carryover effects manifest in food-related CATA studies, and the extent to which they bias responses, remains largely unexplored.

One potential solution to address this bias is to adjust sample evaluation results on the basis of baseline emotional states. Many psychological studies calculate affective change scores after stimulation in order to assess the regulation of emotions by food (Macht & Mueller, 2007), videos (Kuijsters et al., 2016), and music (Groarke & Hogan, 2019). However, only Gallo et al. (2017) applied this approach to CATA data, measuring the baseline once prior to tasting. However, a single baseline measurement cannot capture emotional fluctuations during the session. Furthermore, if emotional carryover does occur across sequential samples, then adjusting with a per-sample baseline, rather than a single per-test baseline, may more effectively isolate the emotional change attributable to each sample.

Therefore, this study had two main objectives. First, from a methodological perspective, the impact of baseline adjustment on emotional responses was examined by comparing three approaches: no baseline adjustment, adjustment using a single per-test baseline, and adjustment using per-sample baselines. Second, from

a psychological process perspective, the carryover effect of emotional states during sequential tasting was investigated, along with its influence on sample evaluations.

2. Material and method

2.1 Participants

Participants were recruited via social media and were students at Jiangnan University. Studies 1 and 2 recruited 129 and 134 participants, respectively. The participants in Study 1 were yogurt consumers, while those in Study 2 were consumers of shrimp sticks (a type of extruded shrimp-flavored snack). The participants were aged 18 to 22 years, with 53-54% females. Before the experiment, they were informed about the products to be tasted and signed the consent form. All of the products were commercially available and safe for consumption. Upon completion of the test, they received snacks as a token of appreciation. The research had been approved by the Ethics Committee of Jiangnan University, Reference number JNU20221201IRB03.

2.2 Sample preparation

Table 1 lists the sample information for Studies 1 and 2. Study 1 included five yogurt samples, all commercially available and selected for their diverse sensory characteristics such as sweetness, viscosity, and flavor. These samples were provided by Classy Kiss Dairy Co., Ltd. (Shenzhen, Guangdong, China). Study 2 involved five shrimp stick samples. The shrimp stick is a type of extruded snack, characterized by its elongated shape and crispy texture. Samples A, B, C, and D were purchased from two different brands (BESTORE Brand and Ai Shang Brand), each differing in terms of their flavor and texture. Sample E was prepared by mixing 0.67 g of Sichuan pepper powder (Lechang Seasoning Co., Ltd., Shandong, China) into 100 g of Sample A, resulting in a slightly pungent sensation. Yogurt is typically associated with nutritional benefits and health (Farag et al., 2021), while shrimp sticks constitute a nostalgic childhood snack for many Chinese consumers.

Therefore, the expectation was that the emotional profiling of these two product categories would differ.

Yogurt samples were removed from the refrigerator 2 hours prior to the experiment to reach room temperature to ensure consistent temperature during tasting. Only yogurt produced within 10 days was used to avoid changes in flavor. Fifteen minutes before the experiment, 20 mL yogurt or 6 g shrimp stick were placed into 30 mL odorless plastic cups (Yuandian Paper Cups Co., Shanghai, China). All samples were labeled with 3-digit random numbers and presented in sequential monadic order using a Williams' Latin square design. For each consumer, the sequence of samples was consistent across the two treatments.

2.3 Experiment design and procedure

A within-subjects experimental design was used to evaluate the influence of baseline emotion questions on emotional responses of samples. In each study, participants were randomly assigned to one of two groups. One group first completed the task without the baseline question, while the other group completed the task with the baseline question. After one month, the participants returned to the lab to complete the other task. The experimental procedure is shown in Fig. 1, the specific procedures are as follow：

The None baseline (None-B) task: For the task with none baseline emotion question, participants were instructed to taste samples and evaluate their overall liking of the sample using a 9-point hedonic scale (ranging from 1= "dislike extremely" to 9= "like extremely"). Subsequently, they answered CATA questions: "Taste the sample, how you feel right now? Please select all that apply". The EsSense25 (Nestrud et al., 2016) with 25 emotion terms was used to test the emotional response and were arranged in a randomized order (King et al., 2013) across two columns. Consumers were given a 2-minute rest period between consecutive samples.

The With Baseline (With-B) task: The design of the with task after receiving the samples was identical to that of the None-B task, except that the baseline emotion questions were asked before receiving each sample (after the 2-minute rest period). The CATA question was: "How do you feel right now? Please select all that apply." The term list for the CATA questionnaire was the same as the one used to measure emotional responses to the samples.

Purified water (Huazhiwei Ltd., Wuxi, Jiangsu, China) and soda biscuits (Aji brand, Weimeng Food Ltd., Dongguan, Guangdong, China) were available for palate cleansing between samples during the course of the experiment. Questionnaires were administered in Chinese, and data were collected via mobile phones using the Wenjuanxing Questionnaire, an online survey platform.

2.4 Data analysis

Data from the two experimental tasks (None-B and With-B) were processed to create four analytical conditions. The None-B condition used sample responses from the task without baseline measurement. The With-B condition used sample responses from the task with baseline measurement, without any adjustment. Two additional conditions were derived by adjusting the With-B data: the Adj. Per-test condition adjusted sample responses using the single baseline collected at the beginning of the test; the Adj. Per-Sample condition adjusted each sample response using its immediately preceding baseline. These four conditions allowed comparison of the effects of baseline measurement and different adjustment strategies on emotional profiling.

2.4.1 Overall liking

Linear mixed models were applied to liking scores from the 9-point hedonic scale. For each study, a primary model was fitted with treatment (None-B, With-B), sample, treatment order, and the treatment × sample interaction as fixed effects, with consumer as a random effect. To examine sample differences within each treatment condition, separate linear mixed models were fitted for the None-B and With-B

conditions, with sample as fixed effect and consumer as random effect. Tukey's test was used for post-hoc comparisons when significant effects were detected.

2.4.3 Emotion term usage

For the None-B and With-B conditions, emotional responses were quantified as the citation frequency of each emotion term.

For the Adj. Per-test and Adj. Per-Sample conditions, the net change from baseline was calculated by subtracting the baseline CATA response from the corresponding sample CATA response for each participant and each emotion term. This subtraction yielded three possible outcomes: a value of +1 indicated that the emotion was selected for the sample but not at baseline (emotion induced); a value of 0 indicated no change in selection status between baseline and sample; and a value of -1 indicated that the emotion was selected at baseline but not for the sample (emotion suppressed). These net change values were then aggregated across participants for each sample and emotion term. Fisher's exact test (Fisher, 1992) was applied to compare term citation frequencies (or net change distributions) between the None-B condition and each of the other three conditions.

2.4.3 Sample differences

For the None-B and With-B conditions, Cochran's Q test was used to identify significant differences among samples in overall term usage and for each individual emotion term (Manoukian, 2022). When significant, pairwise comparisons were conducted using the McNemar test.

For the Adj. Per-test and Adj. Per-Sample conditions, the Kruskal-Wallis test was applied to the net change values for each emotion term, following the procedure proposed by Gallo et al. (2017). The Conover–Iman method was used for multiple comparisons.

2.4.5 Sample and term configurations

For the None-B and With-B conditions, Correspondence Analysis (CA) was performed on the contingency tables of emotion citation frequencies using Hellinger distance (Meyners et al., 2013). For the Adj. Per-test and Adj. Per-Sample conditions, the net change values were aggregated across consumers for each sample and emotion term. A constant was added to eliminate negative values, and CA was performed on the resulting table following the procedure used for Pivot Profile data (Thuillier et al., 2015). The similarity of sample and term configurations across methods was assessed using the RV coefficient on the first two dimensions of the CA solution. Significance was tested by permutation (Robert & Escoufier, 1976).

2.4. 7 Emotional carryover and recovery

To quantify emotional carryover and recovery during the task, the repeatability between adjacent emotion measurements within the With-Baseline condition was assessed. As illustrated in Fig. 2, four types of adjacent pairs were analyzed: Sample-to-Sample (S-S), Baseline-to-Sample (B-S), Sample-to-Baseline (S-B), and Baseline-to-Baseline (B-B).

As suggested by Meyners et al. (2016) for evaluating panel repeatability, the agreement was calculated using Gwet's AC1 statistic. Gwet's AC1 accounts for the expected repeatability by chance and adjusts the raw repeatability rate accordingly (Gwet, 2008). It can be consistently applied across tests with any number of replicates and is less affected by occurrences of very low expected frequencies (Meyners et al., 2016). Gwet's AC1 was calculated using the R code provided by Meyners et al. (2016) with R software (R version 4.3.1, RStudio version 2023.09.1).

2.4.8 Analysis of per-test emotional states

To examine whether consumers' baseline emotional states influence their subsequent CATA responses, participants were first categorized based on their per-test emotional states. Multiple Correspondence Analysis (MCA) was applied to the individual CATA responses from the per-test baseline question. Participants were

grouped based on their coordinates on the first two MCA dimensions, representing emotional valence (positive vs. negative) and arousal (high vs. low), respectively.

Unless otherwise specified, all data organization, management, and statistical analysis were performed using Microsoft Excel (2016) and XLSTAT (2019) software.

3. Results

3.1 Effect of baseline emotion question on concurrent hedonic assessment

Significant differences were observed in overall liking across the samples (Table 2). In Study 1 (yogurt), medium-viscosity, sucrose-added samples (D, E) received the highest liking scores, while the low-viscosity, unsweetened yogurt (A) scored lowest. In Study 2 (shrimp sticks), samples E, A, and C were the most preferred, contrasting with the low overall liking for sample B. Though With-B treatment in Study 2 was found to reduce overall liking score (With-B:5.6 vs. None-B:5.8; p = 0.012), no treatment × sample interaction emerged in either study (p > 0.713), confirming consistent relative preference patterns regardless of baseline assessment.

3.2 Effect of baseline question and baseline adjustment on emotional profiling of samples

3.2.1 Term usage

The inclusion of baseline questions had limited influence on sample question responses. The citation frequencies of terms used by consumers to describe samples showed high similarity between the None-B and With-B tasks (Table 3a), with only one emotion term exhibiting significant differences between the two methodologies (Table 3b).

Baseline adjustment significantly altered the citation frequency of terms used to describe samples. The average number of CATA terms increased significantly (p <0.05) compared to the None-B approach (Table 3a). Specifically, 13-18 terms exhibited statistically significant frequency variations (Table 3b).

3.3.2 Sample discrimination

First, the mean range of citation frequencies for each emotion term across samples varied by method (Table 3c). This metric was derived from the average of per-term maximum-minimum differences. The unadjusted With-B approach slightly reduced this range compared to None-B, whereas the Adj. Per-Sample approach increased it in both studies. Second, regarding the number of emotion terms with significant differences among samples (Table 3d), the unadjusted With-B approach yielded counts similar to None-B. In contrast, both baseline-adjusted methods showed a clear reduction. This led to divergent conclusions for 4 to 9 emotion terms in each study when comparing None-B with the adjusted approaches (Table 3e).

The identity of uniquely discriminating terms provided further insight (Table 4). In Study 1, the unadjusted With-B method uniquely identified terms such as ‘Nostalgic’ and ‘Adventurous’. Across both studies, all three conditions involving baseline measurement consistently identified ‘Disgusted’ as a significantly discriminating term, which the None-B approach did not. Uniquely, in Study 2, the Adj. Per-Sample approach additionally identified ‘Nostalgic’ and ‘Worried’ as discriminating terms. Conversely, several positive or neutral terms that showed significant differences under the None-B approach (e.g., ‘Understanding’, ‘Good’, ‘Loving’) lost significance after baseline adjustment. These results indicate that the inclusion of baseline measurement consistently enhances the detection of differences in negative emotions, with the Adj. Per-Sample method providing the most distinct profile in this regard.

3.2.3 Sample and term configurations

The overall emotional profiles of the samples, as visualized in the symmetric plots (Fig. 3), showed consistent patterns across conditions, with high overall liking samples associated with more positive emotions and specific flavors also influencing emotional responses. High and similar cumulative variance were explained by the first two dimensions across all four methods (range: 88.00% to 95.75%; Tables 3f, 3g). RV coefficients confirmed high similarity in sample configurations (Table 3h, RV > 0.961, $p < 0.0001$). However, the configuration of emotion terms themselves

showed lower congruence, particularly for the Adj. Per-Sample method when compared to None-B (RV as low as 0.751 in Study 2; Table 3i), which corresponds to the results that Adj. approach changed the samples discrimination.

3.2.4 Interpretive shift: From absolute elicitation to directional change

Critically, the interpretation of the citation frequency data differed fundamentally between adjusted and unadjusted methods. The None-B and With-B approaches report the absolute level of emotion elicitation. In contrast, the Adj. Per-Sample and Adj. Per-test approaches report the net change from a baseline state, thereby providing direct information on the direction of emotional change (induction vs. suppression). For example, where the None-B approach indicated yogurt sample A barely induced 'Happy' (1.5% citation), the adjusted approaches revealed it actively reduced happiness (values from -11.2% to -17.2%) relative to baseline (Table S1). Similarly, in Study 2, the adjusted data clarified that three samples suppressed 'Boredom' while one enhanced it.

3.3 Emotional carryover during sequential tasting

3.3.1 Overall carryover effects

To evaluate the changes in emotions during the experimental process, the repeatability between different questions was evaluated by Gwet's AC1 analysis. AC1 values range from -1 to 1. AC1 ≤ 0 indicates no repeatability, while AC1 > 0 indicates repeatability, with higher values reflecting greater consistency.

As shown in Fig. 4, the baseline-to-baseline (B-B) repeatability was substantially below 1 (Study 1: 0.71; Study 2: 0.78), indicating that emotional states did not fully recover to the per-sample level after the 2-minute rest interval.

The sample-to-sample (S-S) agreement was similar between the None-B and With-B methods (Study 1: 0.69 vs. 0.67; Study 2: 0.76 vs. 0.76), confirming that answering baseline questions did not itself alter the inherent emotional interference between consecutive samples. Within this context, the sample-to-baseline (S-B) agreement was higher than the baseline-to-sample (B-S) agreement (e.g., Study 1:

0.74 vs. 0.68; Study 2: 0.78 vs. 0.75). This pattern demonstrates that the emotional impact of a sample persists into the subsequent rest period, exerting a greater influence on the next baseline than the preceding baseline exerts on the sample evaluation. The magnitude of this carryover effect, quantified by the B-S agreement (0.68-0.75), was comparable to the direct sample-to-sample (S-S) agreement (0.67-0.76).

3.3.2 Emotion-specific persistence patterns

Emotion terms with either high or low reproducibility were similar across both studies. Emotions such as ‘Calm’, ‘Mild’, ‘Pleasant’,’ Bored’, and ‘Satisfied’ showed generally low repeatability, indicating significant variation between samples and before and after tasting. Notably, these emotions are associated with low-arousal according to the circumplex model proposed by Jaeger et al. (2018). In contrast, attributes such as "Guilty," "Aggressive," "Nostalgic," "Wild," and "Adventurous" showed high repeatability (AC1 = 0.69·-1.0), with most of these emotions also associated with high-arousal. Additionally, except for "Bored", other negative emotions like "Worried" and "Disgusted" demonstrated higher repeatability. These results suggest that emotions with high-arousal or negative valence are less susceptible to influence by new stimuli and are less prone to fade or recover.

3.4 Effect of per-test emotional states on sample discrimination

To examine how consumers' baseline emotional states influence their subsequent CATA responses, participants were grouped based on their per-test emotional states. This categorization was achieved by applying MCA to individual per-test emotion data (Fig. 5). The resulting first dimension (explaining 43.08–70.28% of inertia) clearly separated consumers reporting negative valence emotions (e.g., Disgusted, Worried) from those reporting positive valence emotions (e.g., Loving, Happy). The second dimension (explaining 5.32–13.56% of inertia) differentiated consumers based on emotional arousal (high vs. low). Participants

were then grouped into high/low valence and high/low arousal categories based on their positions along these two MCA dimensions for subsequent analysis.

Discrimination performance was then compared across valence and arousal groups under different analytical conditions. Before baseline adjustment (With-B), consumers in the negative-valence group consistently identified fewer emotion terms with significant differences between samples than those in the positive-valence group (e.g., Study 1: 19 vs. 23 terms; Study 2: 12 vs. 14 terms; Table 5a). This further supports the notion that negative emotions were less easily altered, making new stimuli more susceptible to the influence of previous samples and per-test emotional states.

Crucially, baseline adjustment moderated this effect. When the Adj. Per-Sample method was applied, the discrimination capacity of the negative-valence group improved relative to the None-B approach in Study 2 (number of discriminating terms increased) and showed a smaller decline in Study 1 (Table 5a, b). Specifically, for the negative-valence group, terms such as ‘Enthusiastic’ (Study 1) and ‘Understanding’ (Study 2) became significant discriminators only after baseline adjustment (Table 5b). In contrast, discrimination in the positive-valence group generally decreased after adjustment.

Next, consumers were divided into two equal groups based on their position along the D2 axis. After grouping, both groups in Study 1 demonstrated similar discriminability by the None-B approach, while in Study 2, high-arousal consumers showed lower discriminability. (Table 5d, e) This further supports the idea that high-arousal emotions are less easily altered. After baseline adjustment, discriminability marginally decreased or remained unchanged for both groups, with a reduction in the number of discriminable attributes by 0–2 (Table 5f).

3.5 Task perceptions

After completing the final test, consumers evaluated the testing task. As shown in the Table 6, consumers perceived the task as easy and not tedious, with no significant differences observed between the two testing methods.

4. Discussion

This research presented here evaluated the methodological and interpretative value of incorporating baseline emotion assessment in food-related emotional profiling. The findings demonstrate a critical two-part conclusion. First, the inclusion of baseline questions itself did not introduce measurable bias or fatigue, thereby establishing the method's practical safety. Second, and more importantly, the adjustment of data against these baselines, particularly using the Per-Sample Adjusted (Adj. Per-Sample) method, fundamentally refined the interpretation of emotional responses. This shifts the paradigm from measuring absolute emotion elicitation to quantifying directional emotional change relative to a per-test state.

4.1 Adding baseline questions does not bias CATA responses but may slightly lower over liking scores

A primary practical concern with adding baseline questions is that lengthy questionnaires might reduce cognitive effort, as suggested by the weak satiation theory (Krosnick, 1991). The unadjusted With-B method yielded sample configurations and term configurations highly similar to the None-B method (RV > 0.96), and showed comparable performance in key metrics such as the number of discriminating terms or sample-to-sample (S-S) agreement. The task perception questionnaire confirmed both methods were perceived as equally easy.

Regarding overall liking, a small reduction was observed with the With-B method in one of the two studies. However, no significant sample × treatment interaction emerged in either study, confirming that baseline questions do not distort the relative liking differences between samples. Previous research has indicated that emotional evaluation of the sample prior to the overall liking evaluation can reduce overall liking scores (King et al., 2013) , whereas conducting the emotional

evaluation after the overall liking evaluation does not introduce hedonic bias (Schouteten et al., 2017). The baseline evaluation appears to exert an effect similar to a sample emotional evaluation. The observation that this hedonic effect occurred in only one study further aligns with evidence that such bias may depend on product category (Jaeger et al., 2020; King et al., 2013). Taken together, while a minor effect on overall liking was observed in one study, it did not compromise the relative differences among samples, and emotional profiles remained unaffected, supporting the practical feasibility of this approach.

4.2 Methodological contributions of baseline adjustment: Revealing directional change and enhancing negative emotion detection

While sample configurations remained stable, indicating consistent relative product relationships, the interpretation of citation frequencies was transformed from absolute elicitation to directional change. Conventional CATA questions (e.g., "How do you feel right now?") capture only the emotional state during tasting (King & Meiselman, 2010; Jaeger, Swaney-Stueve, et al., 2018; Jaeger, Lee, et al., 2018). The Adj. Per-Sample and Adj. Per-test methods extend this paradigm by quantifying whether a sample actively induces or suppresses an emotion relative to the baseline state. This distinction matters because a low citation frequency may indicate not weak elicitation but active suppression. For instance, a sample reported as barely eliciting 'Happy' (1.5%) under the None-B method was shown to actively reduce happiness (-17.2%) after adjustment. Without baseline adjustment, such profiles risk being misinterpreted as emotionally neutral rather than negatively impactful.

Beyond interpretation, baseline measurement and adjustment enhanced the detection of negative emotions. In Study 2, the With-B method identified 'Disgusted' as a discriminator, as well as the two adjusted methods, with the Adj. Per-Sample method uniquely adding 'Worried' (and 'Nostalgic') to the list of discriminators. This demonstrates the unique value of baseline adjustment, particularly the Adj. Per-

Sample approach, in uncovering significant differences in negative emotional dimensions.

Research has shown that individual may regulate negative emotions by eating (Macht et al., 2005), emotional eating behaviors have also been widely reported (Carpio-Arias et al., 2022; Ling & Zahry, 2021). Some food-, like chocolate, only reduced negative emotions induced by film clips, with no effect on induced positive or neutral emotions (Macht & Mueller, 2007). Therefore, understanding the relationship between food and negative emotions is crucial for product development and supporting consumers' emotional health. However, previous research has predominantly focused on positive emotions, as the emotions evoked by food typically have a positive valence (Desmet & Schifferstein, 2008), and the emotion lexicons related to food are predominantly positive (Bhumiratana et al., 2014; King & Meiselman, 2010; Nestrud et al., 2016; Schouteten et al., 2022). This focus may have led to an overlook of how foods influence negative emotions. By revealing directional emotional change and enhancing sensitivity to negative emotional dimensions, baseline adjustment, particularly the Adj. Per-Sample method, addresses this gap.

4.3 Emotional carryover during sequential tasting: Emotion-specific repeatability and its influence on sample discrimination

The Gwet's AC1 analysis revealed that baseline emotional states did not fully recover after the 2-minute rest period, as evidenced by B-B repeatability substantially below. The B-S repeatability was substantial and comparable to the direct S-S repeatability, indicating that the emotional residue from a previous sample influences the next evaluation as much as direct S-S interference. Such carryover effects, where a prior state biases responses to subsequent stimuli, has been observed with various stimuli , such as food (Danner et al., 2016), pictures (Bouhuys et al., 1995), and music (Vuoskoski & Eerola, 2011) Consistent with this behavioral evidence, physiological studies have similarly demonstrated that emotions induced by music or

film stimuli require more than 2 minutes to fully return to baseline levels (Kuijsters et al., 2016; Ribeiro et al., 2019).

Critically, not all emotions showed the same degree of repeatability. Negative and high-arousal emotions (e.g., 'Guilty', 'Aggressive', 'Disgusted') showed the highest repeatability across all measurement pairs, indicating they are less altered by new stimuli or rest. This phenomenon may relate to the concept of a refractory period in emotional experience. Negrão et al. (2022) found that consumers with more negative emotions were less likely to change their emotional state after watching a movie and explained this phenomenon using the concept of the refractory period in emotional experience (Ekman & Cordaro, 2011). Individuals experiencing strong emotions, particularly negative ones, may enter a refractory period, a state in which they are less affected by external stimuli that could modulate their current emotional state. During this period, the brain filters incoming information, allowing only those stimuli with emotional valence compatible with the current state (i.e., negative) to be processed (Kuppens et al., 2010).

This differential repeatability provides a unifying explanation for several interrelated findings in the present study. First, it accounts for the observation that consumers with negative per-test emotional states exhibited weaker sample discrimination when data were not adjusted for baseline: their persistent emotional background reduced sensitivity to sample-induced changes. Second, it explains why the Adj. Per-Sample baseline adjustment specifically improved discrimination for these consumers and enhanced the detection of negative emotions reported in Section 4.2: by subtracting the per-sample baseline, the method isolates the change attributable to the sample itself, effectively correcting for the bias introduced by persistent emotional states.

This differential repeatability also offers a mechanistic explanation for the task-related effect observed by King et al. (2013). They reported that with increasing sample numbers, responses become less sample-specific and more influenced by

the task itself, leading them to suggest limiting emotion studies to just two samples. Notably, they found that emotions such as calm, interested, quiet, and satisfied decreased progressively across serving positions, while affectionate increased. Our data suggest this pattern may be driven, in part, by the accumulation of persistent high-arousal and negative emotional states across sequential samples, which could reduce responsiveness to new stimuli.

4 Limitations and suggestions for future research

Baseline adjustment reduced overall sample discrimination, a primary objective in most conventional sensory evaluation contexts. While the method contributed to the detection of negative emotions and directional emotional change, this limits its suitability as a universal replacement for standard CATA practices. The approach is therefore best positioned as a targeted supplement for research specifically focused on emotional modulation or negative emotion regulation.

Additionally, the limited impact on overall sample discrimination may stem from the method's inability to capture intensity increases when an emotion is already present at baseline. Future studies could incorporate intensity-based approaches such as Rate-All-That-Apply (RATA) to better detect such variations.

The generalizability of the findings remains to be established. Both studies usemployed fermented or marine products within controlled laboratory settings. Future research should extend this paradigm to a wider range of product categories and consumption contexts to determine whether the value of baseline adjustment varies across different conditions.

5. Conclusions

This study examined whether incorporating per-sample baseline emotion measurement in CATA profiling provides methodological value beyond conventional approaches. Adding baseline questions did not compromise CATA responses or increase perceived task difficulty, though a minor reduction in overall liking was observed in one study. Building on this foundation, two key advantages of baseline

adjustment were identified. First, it shifts the interpretation of emotional responses from absolute elicitation to directional change, revealing that foods not only induce emotions but can also actively suppress them. Second, while baseline adjustment modestly reduces the overall number of discriminating terms, there is an indication that it may enhance the detection of certain negative emotions (e.g., ‘Disgusted’, ‘Worried’) and improve sample discrimination for consumers with negative per-test emotional state.

The study also confirmed that emotional states do not fully recover during a brief rest and that negative, high-arousal emotions exhibit greater persistence across sequential samples. This differential persistence may help explain why consumers with negative per-test emotional states showed reduced discrimination under conventional methods, and why per-sample baseline adjustment appeared to mitigate this effect. By isolating the emotional change attributable to the sample itself, baseline adjustment offers a potential means to account for emotional carryover in sequential testing.

In light of these findings, the inclusion of baseline emotion questions prior to each sample is recommended not as a universal replacement for conventional practices, but as a targeted methodological supplement. It is particularly valuable when the research objective involves elucidating the directional emotional impact of a product or evaluating its potential to alleviate negative emotional states. When these questions are not central to the research aim, the conventional None-B approach remains both efficient and valid.

Acknowledgements

This research was supported by the International S&T Innovation Cooperation Key Project (2024YFE0112700) and the program of “Collaborative Innovation Center of Food Safety and Quality Control in Jiangsu Province”, China

REFERENCES

Adolphs, R., & Andler, D. (2018). Investigating Emotions as Functional States Distinct From Feelings. *Emotion Review*, *10*(3), 191–201. https://doi.org/10.1177/1754073918765662

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

Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, *33*(4), 215–226. https://doi.org/10.1016/0165-0327(94)00092-N

Carpio-Arias, T. V., Solís Manzano, A. M., Sandoval, V., Vinueza-Veloz, A. F., Rodríguez Betancourt, A., Betancourt Ortíz, S. L., & Vinueza-Veloz, M. F. (2022). Relationship between perceived stress and emotional eating. A cross sectional study. *Clinical Nutrition ESPEN*, *49*, 314–318. https://doi.org/10.1016/j.clnesp.2022.03.030

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. https://doi.org/10.1371/journal.pone.0115388

Danner, L., Ristic, R., Johnson, T. E., Meiselman, H. L., Hoek, A. C., Jeffery, D. W., & Bastian, S. E. P. (2016). Context and wine quality effects on consumers' mood, emotions, liking and willingness to pay for Australian

559 Shiraz wines. *Food Research International*, *89*, 254–265. https://doi.org/
560 10.1016/j.foodres.2016.08.006
561 Desmet, P. M. A., & Schifferstein, H. N. J. (2008). Sources of positive and
562 negative emotions in food experience. *Appetite*, *50*(2–3), 290–301.
563 https://doi.org/10.1016/j.appet.2007.08.003
564 Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic.
565 *Emotion Review*, *3*(4), 364–370.
566 https://doi.org/10.1177/1754073911410740
567 Farag, M. A., Saleh, H. A., & Ahmady, S. E. (2021). Dissecting Yogurt: The
568 Impact of Milk Types, Probiotics, and Selected Additives on Yogurt
569 Quality. *Food Reviews International*, *38*(sup1), 634–650. https://doi.org/
570 10.1080/87559129.2021.1877301
571 Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz & N.
572 L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and*
573 *Distribution* (pp. 66–70). Springer New York.
574 https://doi.org/10.1007/978-1-4612-4380-9_6
575 Gallo, K. E., Swaney-Stueve, M., & Chambers, D. H. (2017). Comparing visual
576 food images versus actual food when measuring emotional response of
577 children. *Journal of Sensory Studies*, *32*(3), e12267.
578 https://doi.org/10.1111/joss.12267
579 Gomez, P., Zimmermann, P. G., Guttormsen Schär, S., & Danuser, B. (2009).
580 Valence Lasts Longer than Arousal. *Journal of Psychophysiology*, *23*(1),
581 7–17. https://doi.org/10.1027/0269-8803.23.1.7

582 Groarke, J. M., & Hogan, M. J. (2019). Listening to self-chosen music regulates
583 induced negative affect for both younger and older adults. *PloS One*,
584 *14*(6), e0218017. https://doi.org/10.1371/journal.pone.0218017
585 Gunaratne, T. M., Gonzalez Viejo, C., Fuentes, S., Torrico, D. D., Gunaratne, N.
586 M., Ashman, H., & Dunshea, F. R. (2019). Development of emotion
587 lexicons to describe chocolate using the Check-All-That-Apply (CATA)
588 methodology across Asian and Western groups. *Food Research*
589 *International*, *115*, 526–534.
590 https://doi.org/10.1016/j.foodres.2018.10.001
591 Gupta, M., Torrico, D. D., Hepworth, G., Gras, S. L., Ong, L., Cottrell, J. J., &
592 Dunshea, F. R. (2021). Differences in Hedonic Responses, Facial
593 Expressions and Self-Reported Emotions of Consumers Using
594 Commercial Yogurts: A Cross-Cultural Study. *Foods*, *10*(6), 1237. https://
595 doi.org/10.3390/foods10061237
596 Gutjar, S., De Graaf, C., Kooijman, V., De Wijk, R. A., Nys, A., Ter Horst, G. J., &
597 Jager, G. (2015). The role of emotions in food choice and liking. *Food*
598 *Research International*, *76*, 216–223.
599 https://doi.org/10.1016/j.foodres.2014.12.022
600 Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the
601 presence of high agreement. *British Journal of Mathematical and*
602 *Statistical Psychology*, *61*(1), 29–48.
603 https://doi.org/10.1348/000711006X126600
604 Jaeger, S. R., Jin, D., Roigard, C. M., Le Blond, M., & Ares, G. (2020). Risk of

hedonic bias in sensory co-elicitations: Comparison of CATA questions and applicability ratings. *Journal of Sensory Studies*, *35*(5), 1–10. https://doi.org/10.1111/joss.12601

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. https://doi.org/10.1016/j.foodqual.2018.04.001

Jaeger, S. R., Spinelli, S., Ares, G., & Monteleone, E. (2018). Linking product-elicited emotional associations and sensory perceptions through a circumplex model based on valence and arousal: Five consumer studies. *Food Research International*, *109*, 626–640. https://doi.org/10.1016/j.foodres.2018.04.063

Jaeger, S. R., Swaney-Stueve, M., Chheang, S. L., Hunter, D. C., Pineau, B., & Ares, G. (2018). An assessment of the CATA-variant of the EsSense Profile®. *Food Quality and Preference*, *68*, 360–370. https://doi.org/10.1016/j.foodqual.2018.04.005

Kaneko, D., Brouwer, A.-M., Hogervorst, M., Toet, A., Kallen, V., & 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*, 558172. https://doi.org/10.3389/fpsyg.2020.558172

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

628 *Preference*, *21*(2), 168–177.
629 https://doi.org/10.1016/j.foodqual.2009.02.005
630 King, S. C., Meiselman, H. L., & Thomas Carr, B. (2013). Measuring emotions
631 associated with foods: Important elements of questionnaire and test
632 design. *Food Quality and Preference*, *28*(1), 8–16.
633 https://doi.org/10.1016/j.foodqual.2012.08.007
634 Krosnick, J. A. (1991). Satisficing among reluctant respondents. *Applied*
635 *Cognitive Psychology*, *5*(3), 213–236.
636 https://doi.org/10.1002/acp.2350050305
637 Kuijsters, A., Redi, J., De Ruyter, B., & Heynderickx, I. (2016). Inducing Sadness
638 and Anxiousness through Visual Media: Measurement Techniques and
639 Persistence. *Frontiers in Psychology*, *7*.
640 https://doi.org/10.3389/fpsyg.2016.01141
641 Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional Inertia and
642 Psychological Maladjustment. *Psychological Science*, *21*(7), 984–991.
643 https://doi.org/10.1177/0956797610372634
644 Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles*
645 *and Practices*. Springer New York. https://doi.org/10.1007/978-1-4419-
646 6488-5
647 Ling, J., & Zahry, N. R. (2021). Relationships among perceived stress,
648 emotional eating, and dietary intake in college students: Eating self-
649 regulation as a mediator. *Appetite*, *163*, 105215.
650 https://doi.org/10.1016/j.appet.2021.105215

Macht, M., Haupt, C., & Ellgring, H. (2005). The perceived function of eating is changed during examination stress: A field study. *Eating Behaviors*, *6*(2), 109–112. https://doi.org/10.1016/j.eatbeh.2004.09.001

Macht, M., & Mueller, J. (2007). Immediate effects of chocolate on experimentally induced mood states. *Appetite*, *49*(3), 667–674. https://doi.org/10.1016/j.appet.2007.05.004

Manoukian, E. B. (2022). *Mathematical nonparametric statistics* (First edition). Routledge.

Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference*, *30*(2), 309–319. https://doi.org/10.1016/j.foodqual.2013.06.010

Meyners, M., Castura, J. C., & Worch, T. (2016). Statistical evaluation of panel repeatability in Check-All-That-Apply questions. *Food Quality and Preference*, *49*, 197–204. https://doi.org/10.1016/j.foodqual.2015.12.010

Negrão, J. G., Bazán, P. R., De Azevedo Neto, R. M., Lacerda, S. S., Ekman, E., & Kozasa, E. H. (2022). Baseline emotional state influences on the response to animated short films: A randomized online experiment. *Frontiers in Psychology*, *13*, 1009429. https://doi.org/10.3389/fpsyg.2022.1009429

Nestrud, M. A., Meiselman, H. L., King, S. C., Lesher, L. L., & Cardello, A. V. (2016). Development of EsSense25, a shorter version of the EsSense Profile®. *Food Quality and Preference*, *48*, 107–117.

https://doi.org/10.1016/j.foodqual.2015.08.005

Ribeiro, F. S., Santos, F. H., Albuquerque, P. B., & Oliveira-Silva, P. (2019). Emotional Induction Through Music: Measuring Cardiac and Electrodermal Responses of Emotional States and Their Persistence. *Frontiers in Psychology*, *10*, 451. https://doi.org/10.3389/fpsyg.2019.00451

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. https://doi.org/10.1016/j.foodres.2015.11.001

Schouteten, J. J., Gellynck, X., De Bourdeaudhuij, I., Sas, B., Bredie, W. L. P., Perez-Cueto, F. J. A., & De Steur, H. (2017). Comparison of response formats and concurrent hedonic measures for optimal use of the EmoSensory® Wheel. *Food Research International*, *93*, 33–42. https://doi.org/10.1016/j.foodres.2016.12.015

Schouteten, J. J., Verwaeren, J., Rini, L., & Almli, V. L. (2022). Comparing a product-specific versus a general emoji list to measure consumers' emotional associations with chocolate and predict food choice. *Food Research International*, *157*, 111299. https://doi.org/10.1016/j.foodres.2022.111299

Souza-Coutinho, M., Brasil, R., Souza, C., Sousa, P., & Malfeito-Ferreira, M. (2020). Consumers Associate High-Quality (Fine) Wines with

Complexity, Persistence, and Unpleasant Emotional Responses. *Foods*, *9*(4), 452. https://doi.org/10.3390/foods9040452

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. https://doi.org/10.1016/j.foodqual.2014.12.009

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. https://doi.org/10.1016/j.foodqual.2010.04.011

Thuillier, B., Valentin, D., Marchal, R., & Dacremont, C. (2015). Pivot profile: A new descriptive method based on free description. *Food Quality and Preference*, *42*, 66–77. https://doi.org/10.1016/j.foodqual.2015.01.012

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex*, *47*(9), 1099–1106. https://doi.org/10.1016/j.cortex.2011.04.011

713 **Table1. Overview of the sample information**

Study ID	Product category	Sample Information
Study 1	Yogurt	A: No added sweetener, low viscosity
		B: Added sucrose, low viscosity
		C: Added low and no calorie sweeteners, low viscosity
		D: Added sucrose, medium viscosity
		E: Added sucrose, medium viscosity, strawberry flavor
Study 2	Shrimp stick	A: Shrimp flavor, BESTORE Brand
		B: Crab flavor, Ai Shang Brand
		C: Spicy crawfish flavor, Ai Shang Brand
		D: Shrimp flavor, Ai Shang Brand
		E: Sample A with Sichuan pepper powder added

714

715

716

Table 2. Mean overall liking scores for each sample under None-B and With-B conditions, with results from linear mixed models (treatment, sample, treatment × sample) and Tukey's post-hoc comparisons.

Study	Sample	None-B	With-B	*p* value (sample)	*p* value (method)	*p* value (sample *method)
Study 1	A	2.3[a]	2.3[a]	**< 0.0001**	0.144	0.713
	B	5.3[b]	5.5[b]			
	C	5.1[b]	5.2[b]			
	D	6.3[c]	6.5[c]			
	E	6.8[c]	6.7[c]			
	Average	5.2	5.3			
Study 2	A	6.3[c]	6.2[c]	**< 0.0001**	**0.012**	0.986
	B	4.9[a]	4.6[a]			
	C	6.0[c]	5.8[c]			
	D	5.5[b]	5.3[b]			
	E	6.4[c]	6.2[c]			
	Average	5.8	5.6			

[abc] Mean value with different letters in a column within on study are significantly different at $p < 0.05$

Table 3. Summary of results for the comparison of emotion characterizations by consumers obtained from None-B, With-B, Adj. Per-sample and Adj. Per-test.

	Study1				Study2			
	None-B	With-B	Adj. Per-sample	Adj. Per-test	None-B	With-B	Adj. Per-sample	Adj. Per-test
Term Usage								
a. Average nNumber of terms used to describe samples (or change number from baseline)	3.7a	3.5a	5.7b	5.9b	3.2a	3.2a	4.1b	4.6b
b. Number of CATA terms where the approach differ significantly in frequency of use at the aggregate level ($p < 0.05$) compared to None-B	-	1	18	17	-	1	13	13
Sample differences								
c. Average range of emotion term citation frequencies between samples (%)	25.5	23.2	27.1	23.2	13.2	12.2	14.3	12.2
d. Number of CATA terms with significant differences among samples	23	24	19	18	16	15	14	12
e. Number of CATA terms where different conclusions regarding similarities and differences among samples were drawn for the method	-	3	4	5	-	4	9	7

Sample configurations								
f. Percentage of variance explained by Dimensions 1 and 2	95.8%	94.9%	95.5%	95.5%	90.1%	89.0%	88.0%	92.6%
g. Percentage of variance explained by Dimension 1	83.3%	82.4%	75.7%	76.3%	76.2%	72.5%	68.2%	80.5%
h. RV coefficient between sample configurations (vs. None-B)	-	0.993	0.962	0.974	-	0.981	0.969	0.961
i. RV coefficient between term configurations (vs. None-B)	-	0.919	0.910	0.922	-	0.830	0.751	0.801

None-B: emotional response of samples without baseline emotion measurement;

With-B: emotional response of samples with baseline emotion measurement;

Adj. Per-sample: The frequency of each emotion term for a given sample was calculated as the difference between the post-tasting response and the baseline response collected immediately before tasting that same sample;

Adj. Per-test: The frequency of each emotion term for all samples was calculated as the difference between the post-tasting responses and the single baseline response collected at the very beginning of the entire tasting session.

Table 4. Unique discriminative emotion terms compared to the None-B method.

Study	Method	Terms Significant in None-B but NOT in This Method	Terms Significant in This Method but NOT in None-B
1	With-B	Free	Nostalgic, Adventurous
	Adj. Per-sample	Understanding, Good Natured, Free, Active	—
	Adj. Per-test	Bored, Understanding, Good Natured, Free, Active	—
2	With-B	Understanding, Good, Loving	Disgusted
	Adj. Per-Sample	Understanding, Happy, Mild, Tame, Good, Loving	Disgusted, Nostalgic, Worried
	Adj. Per-test	Understanding, Mild, Tame, Good, Calm, Loving	Disgusted

734 **Table 5. Comparison of sample discrimination results between With-B and Adj. Per-Sample methods for consumers grouped by their per-**

735 **test emotional state.**

Group method	Sample differences	Group	Study 1		Study 2	
			With-B	Adj. Per-Sample	With-B	Adj. Per-Sample
Valence	a. Number of CATA terms with significant differences among samples	Positive	23	17	14	7
		Negative	19	15	12	13
	b. CATA terms with significant differences among sample detected uniquely by the method	Positive	Bored, Good, Active Natured, Free, Guilty, Adventurous		Pleasant, Happy, Joyful Interested, Tame, Enthusiastic, Aggressive, Calm,	Free
		Negative	Nostalgic, Mild, Tame, Secure, Aggressive	Enthusiastic	Mild	Understanding, Interested
	c. Number of CATA terms where different conclusions regarding similarities and differences among samples were drawn for the method	Positive	6		9	
		Negative	6		4	
Arousal	d. Number of CATA terms with significant differences among samples	High	21	21	7	5
		Low	21	20	13	12
	e. CATA terms where different conclusions regarding similarities and differences among samples were drawn for the method	High			Active, Calm	
		Low	Adventurous		Happy, Tame, Calm, Joyful	Nostalgic, Wild, Secure
	f. Number of CATA terms where	High	0		2	

	different conclusions regarding similarities and differences among samples were drawn for the method	Low	1	7

None-B: emotional response of samples without baseline emotion measurement; WB-PreS: emotional response of samples adjusted by baseline emotion before each sample.

738 **Table 6. Summary of results for the comparison of task perceptions for None-B**
739 **and With-B task. Values shown are means with standard deviations between**
740 **brackets.**

Study	Methodology	It was easy to answer the questions	It was tedious to answer the questions
Study 1	With-B	4.8(1.4)	3.5(1.4)
	None-B	4.4(1.3)	3.9(1.6)
	p value	0.191	0.130
Study 2	With-B	4.6(1.3)	3.9(1.6)
	None-B	4.7(1.3)	3.7(1.5)
	p value	0.843	0.424

741 None-B: emotional response of samples without baseline emotion measurement; With-B:
742 emotional response of samples with baseline emotion measurement.

Figure Captions

Fig. 1. Overview of the experimental procedure for participants in the research.

Fig. 2. Illustration of the four types of adjacent emotion measurement pairs analyzed within the With-B condition: sample-to-sample (S-S), baseline-to-sample (B-S), sample-to-baseline (S-B), and baseline-to-baseline (B-B).

Fig. 3. MFA emotion plots from the None-B, With-B, WB-PreT, and WB-PreS approaches. (a) Study 1, (b) Study 2, and sample plots from the four approaches: (c) Study 1, (d) Study 2. None-B: emotional response of samples without baseline emotion measurement; WB: emotional response of samples with baseline emotion measurement; WB-PreT: emotional response of samples adjusted by baseline emotion before the whole testing; WB-PreS: emotional response of samples adjusted by baseline emotion before each sample.

Fig. 4. AC1 value for adjacent emotional response in baseline to baseline (B-B), sample to baseline (S-B), baseline to sample (B-S) and sample to sample (S-S) in Study 1 and

Fig. 5. Multivariate Correspondence Analysis (MCA) of individual per-test emotional states in Study 1 (a) and Study 2 (b).

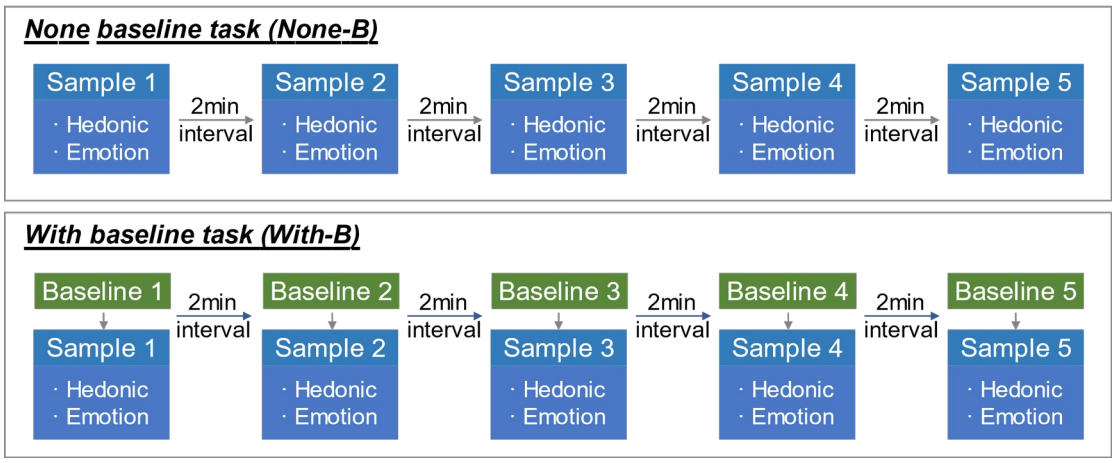


Fig. 1. Overview of the experimental procedure for participants in the research.

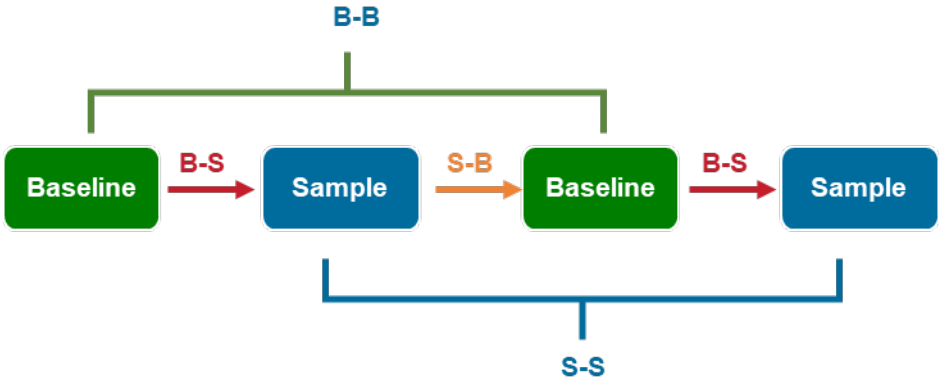


Fig. 2. Illustration of the four types of adjacent emotion measurement pairs analyzed within the With-B condition: sample-to-sample (S-S), baseline-to-sample (B-S), sample-to-baseline (S-B), and baseline-to-baseline (B-B).

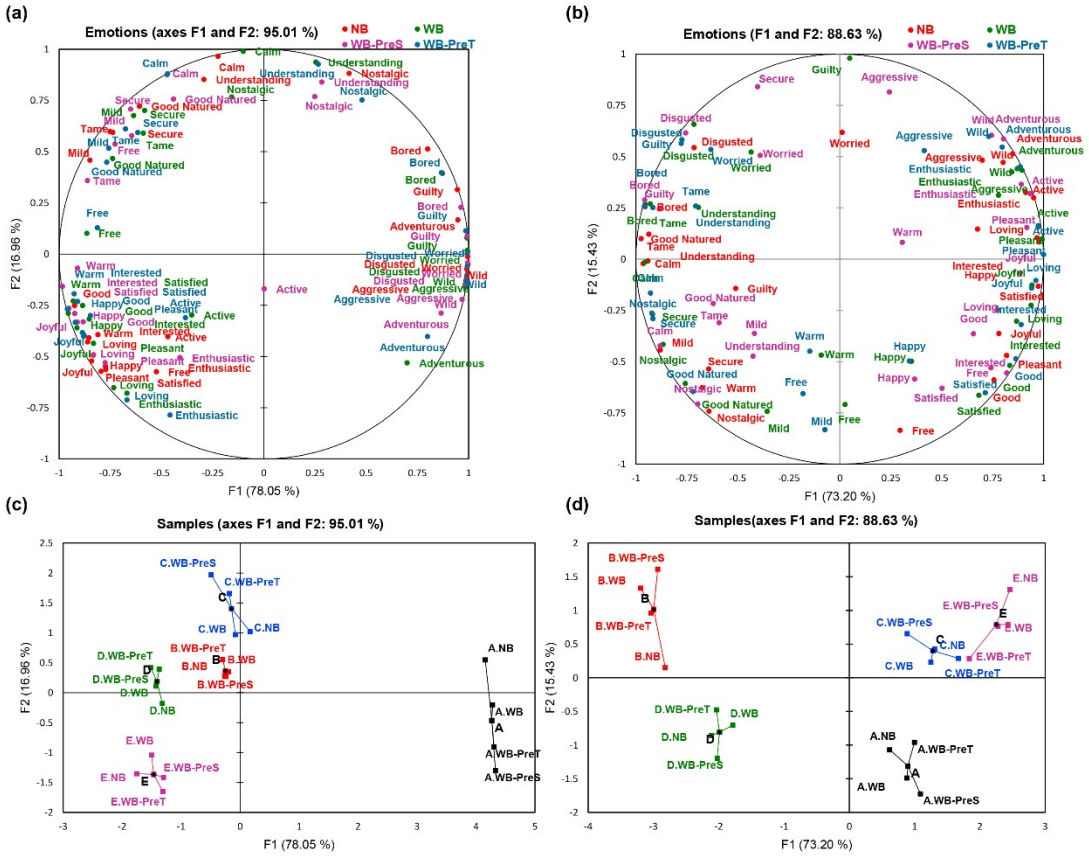
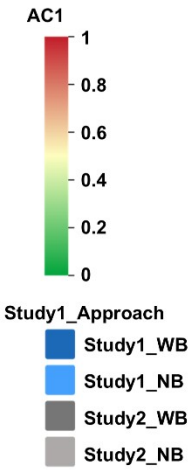


Fig. 3. MFA emotion plots from the None-B, With-B, WB-PreT, and WB-PreS approaches. (a) Study 1, (b) Study 2, and sample plots from the four approaches: (c) Study 1, (d) Study 2. None-B: emotional response of samples without baseline emotion measurement; WB: emotional response of samples with baseline emotion measurement; WB-PreT: emotional response of samples adjusted by baseline emotion before the whole testing; WB-PreS: emotional response of samples adjusted by baseline emotion before each sample.

	Study1					Study2				
	B-B	S-B	B-S	S-S	S-S	B-B	S-B	B-S	S-S	S-S
Guilty	0.96	0.94	0.94	0.90	0.87	0.98	1.00	0.99	0.99	0.99
Aggressive	0.96	0.93	0.91	0.87	0.83	0.99	0.95	0.95	0.92	0.94
Nostalgic	0.88	0.88	0.87	0.89	0.81	0.93	0.89	0.88	0.84	0.82
Wild	0.92	0.86	0.83	0.74	0.69	0.97	0.94	0.93	0.90	0.90
Adventurous	0.87	0.85	0.83	0.86	0.85	0.92	0.89	0.90	0.85	0.85
Loving	0.92	0.88	0.83	0.75	0.74	0.95	0.93	0.91	0.87	0.86
Disgusted	0.89	0.87	0.77	0.67	0.61	0.99	0.98	0.96	0.94	0.96
Worried	0.81	0.84	0.75	0.74	0.67	0.92	0.94	0.92	0.91	0.92
Enthusiastic	0.85	0.87	0.84	0.82	0.77	0.86	0.83	0.79	0.78	0.76
Free	0.69	0.76	0.72	0.80	0.79	0.77	0.79	0.80	0.83	0.82
Warm	0.76	0.81	0.73	0.76	0.77	0.81	0.80	0.79	0.73	0.79
Happy	0.74	0.77	0.69	0.67	0.61	0.84	0.79	0.79	0.76	0.73
Active	0.76	0.77	0.70	0.72	0.67	0.78	0.77	0.75	0.72	0.71
Good Natured	0.65	0.74	0.70	0.77	0.73	0.67	0.73	0.69	0.78	0.77
Secure	0.59	0.68	0.62	0.74	0.65	0.70	0.72	0.75	0.81	0.79
Underatanding	0.65	0.67	0.65	0.64	0.61	0.75	0.77	0.75	0.73	0.82
Tame	0.64	0.64	0.64	0.65	0.64	0.74	0.74	0.76	0.82	0.75
Juyful	0.71	0.72	0.60	0.53	0.66	0.72	0.71	0.70	0.71	0.72
Interested	0.66	0.68	0.67	0.64	0.67	0.66	0.70	0.64	0.65	0.67
Good	0.70	0.67	0.59	0.49	0.58	0.75	0.72	0.72	0.65	0.63
Satisfied	0.64	0.66	0.58	0.49	0.58	0.74	0.69	0.64	0.59	0.62
Bored	0.49	0.69	0.51	0.66	0.61	0.61	0.69	0.52	0.57	0.66
Pleasant	0.53	0.65	0.46	0.46	0.41	0.70	0.68	0.63	0.66	0.58
Mild	0.38	0.48	0.44	0.59	0.54	0.42	0.53	0.51	0.70	0.67
Calm	0.19	0.26	0.11	0.50	0.44	0.29	0.22	0.20	0.37	0.39
Average	0.71	0.74	0.68	0.69	0.67	0.78	0.78	0.75	0.76	0.76

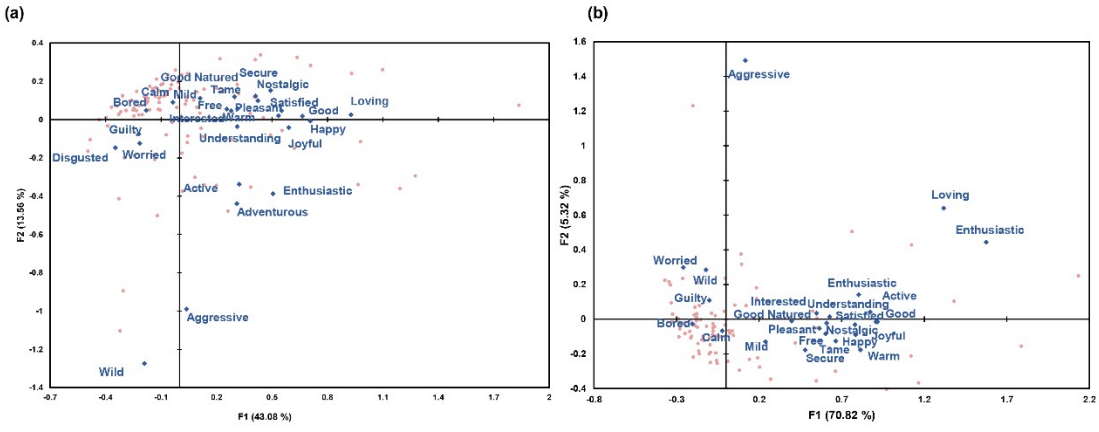


780

781 Fig. 4. AC1 value for adjacent emotional response in baseline to baseline (B-B), sample
782 to baseline (S-B), baseline to sample (B-S) and sample to sample (S-S) in Study 1 and
783 2.

784

785



786

787 Fig. 5. Multivariate Correspondence Analysis (MCA) of individual per-test emotional
788 states in Study 1 (a) and Study 2 (b).