sp open
swiss psychology open

Aesthetics of Taste in Evaluations of User Experience

RESEARCH

ANDREAS SONDEREGGER
JUERGEN BAUMGARTNER
JUERGEN SAUER

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

The important role aesthetics plays in human-computer interaction has been widely addressed in scientific research. While most of this research focused on visual aesthetics, other sensory modalities have received considerably less attention. The present article reports three experiments (N = 40; N = 60; N = 38) adopting differing experimental designs to address the influence of aesthetics of taste in consumer product evaluation. Using a fully automatic coffee machine as a model product, we examined the impact of aesthetics of taste on the evaluation of user experience, including measures of perceived usability, performance and affect. In the three experiments, the taste of coffee was manipulated together with system usability as an important factor of user experience. The findings showed that aesthetics of taste had little effect on user evaluations of the coffee machine. The results were in stark contrast to the medium to large effects of visual aesthetics found in numerous previous studies but they were similar to the findings of some research on auditory and tactile aesthetics. Overall, the present work provides further evidence for differences between the influence of non-visual and visual aesthetics in evaluations of user experience. Moreover, the three studies show that in the field of experimental research, the use of contrast effects for the experimental manipulation of aesthetics is a promising method, indicating important insights for future experimental research in the domain of design aesthetics.

RESEARCH HIGHLIGHTS

- The three studies were the first to examine the influence of aesthetics of taste on user experience.
- In contrast to visual aesthetics, aesthetics of taste did only marginally influence measures of user experience.
- The work suggests potential differences between non-visual and visual aesthetics.
- The findings indicate that for experimental research in aesthetics the use of the contrast effect is promising.

CORRESPONDING AUTHOR:
Andreas Sonderegger
FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
andreas.sonderegger@fhnw.ch

KEYWORDS:
gustatory aesthetics; user experience; coffee machine; experimental aesthetics

TO CITE THIS ARTICLE:
Sonderegger, A., Baumgartner, J., & Sauer, J. (2026). Aesthetics of Taste in Evaluations of User Experience. *Swiss Psychology Open,* 6(1): 8, pp. 1–23. DOI: https://doi.org/10.5334/spo.94

INTRODUCTION

Although the sense of taste has so far been rather neglected in the field of human-computer interaction (HCI), it represents an important sensory modality in various areas of the consumer goods industry, such as for the development of machines that produce beverages and food (e.g. coffee machines, 24-hour pizza machines, food dispensers; cf. Cameron et al., 2020; Murer et al., 2013; Obrist et al., 2014). In addition, current trends in 3D food printing technology point to the increasing practical importance of taste in the field of HCI (Gayler et al., 2019; Gayler & Sas, 2017). Regarding the design of such technology, rather little is known about how user experience is influenced by different aspects of the usage situation. Until now, considerable work has been carried out demonstrating that visual aesthetics has an important influence on user experience (UX), an effect most often referred to as the 'what is beautiful is usable'-effect (Tractinsky et al., 2000), based on a similar effect previously found in social psychology ('what is beautiful is good'-stereotype, WIBIG; Dion et al., 1972).

Most of this work dealt with the effect of visual aesthetics. Although vision may be considered the predominant sensory modality for product experience and evaluation, other senses such as taste and sound may also play an important role (Schifferstein, 2006). This applies particularly to the user interaction with products providing edible or drinkable outcomes. With regard to such technology, subjective evaluations of the outcomes of the user-device interaction (e.g., coffee as outcome of the interaction of a user with a coffee machine) could have an important influence in product evaluations (Hornbæk, 2006). In this regard, there is a gap in the theoretical framework regarding whether people can distinguish between aesthetics of taste and UX. Furthermore, this highlights the practical relevance of the research questions as UX practitioners might need to understand how strongly the experience of test users is influenced (or biased) by the taste of the outcome product.

UX EVALUATION AND THE INFLUENCE OF AESTHETICS

In the context of this piece of research, UX is considered a holistic umbrella construct encompassing all aspects a user experiences when interacting with a system (cf. Sauer et al., 2020, see also Dewey, 2005; Wright et al., 2003). Because these experiential consequences of the interaction are highly complex (i.e. consisting of attitudes, affect, beliefs, and behavioural measures, etc.), UX should not be assessed as a one-dimensional concept (i.e. providing a single score). Instead, it should be represented by measuring the most important UX components, including perceived usability, affect and emotions, efficiency and effectiveness (Sonderegger et al., 2019, see also; Sauro & Lewis, 2012; Law & van Schaik, 2010).

Usability has been defined by ISO FDIS 9241-210 as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-210, 2019, p. 3). The concept is most often assessed subjectively by means of usability scales (see Assila et al., 2016 for an overview of published scales), of which the System Usability Scale (Brooke, 1996) is considered one of the very well established and most often cited questionnaires in the usability domain (Lewis, 2018). Furthermore, objective behavioural outcomes are assessed as indicators of the efficiency and effectiveness components of the usability definition. Emotions are generally considered a multifaceted concept, representing a process or episode which is triggered by a specific event (Sander, 2013). Most often, five components of an emotion episode are differentiated, such as 1) subjective appraisal (cognitive component), 2) physiological reactions (neurophysiological component), 3) action tendencies (motivational component), 4) facial and vocal expressions (motor expression component) and 5) emotional experience (subjective feeling component) (Scherer, 2005). Each of these components can be measured with appropriate tools and instruments, in the context of UX evaluations, emotions are most often assessed subjectively via questionnaires. In assessing the subjective component of emotions, a distinction can be made between a dimensional approach and an approach focusing on the assessment of discrete emotions. As an example for the dimensional approach, the circumplex model of affect differentiates between valence, arousal and dominance as the three dimensions of an emotion episode (Russell, 1980). Although this framework proposes dominance as a third affective dimension, assessment is often restricted to the dimensions of valence and arousal. This is because these two dimensions have been found to capture most variance in subjective affective experience, whereas dominance often provides only limited additional explanatory value. Moreover, dominance is frequently interpreted as reflecting cognitive appraisals of control or agency rather than a core affective experience itself (Russell & Barrett, 1999). Subjective assessment of discrete emotions focuses on measuring specific emotions (e.g., happiness, anger, fear, sadness). Common self-report instruments include Discrete Emotions Questionnaire (DEQ; Harmon-Jones et al., 2016), Product Emotion Measure (PrEmo; Desmet et al., 2000), Geneva Emotion Wheel (Scherer et al., 2013), or Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). Due to its robust psychometric properties, extensive cross-cultural validation, and ease of administration, the PANAS has

become one of the most frequently used instruments for measuring subjective emotional experience.

Overall, there are a considerable number of studies that have examined the influence of aesthetics[1] in product evaluation. Research on the effects of aesthetics on perceived usability has provided a highly consistent picture with higher aesthetics resulting in better usability ratings of the product (e.g., Ben-Bassat et al., 2006; Sonderegger & Sauer, 2010; Thielsch et al., 2014; Thüring & Mahlke, 2007; Tractinsky, 1997). This effect is considered rather robust with regard to cultural influences since work in several countries has shown a highly similar picture, including Japan (Kurosu & Kashimura, 1995), Israel (Tractinsky, 1997), Malaysia (Yusof et al., 2010), Korea (Seo et al., 2016), United Kingdom (De Angeli et al., 2006), Germany (Thielsch et al., 2014) and Switzerland (Sonderegger, Sauer, et al., 2014).

The 'halo effect' has been cited most prominently to explain the positive association between the product's aesthetic appeal and its perceived usability (e.g., Tractinsky et al., 2000), though alternative theoretical explanations for this effect were also suggested (e.g., Hassenzahl & Monk, 2010, De Angeli et al., 2006; for a complete overview see Thielsch et al., 2015). The halo effect refers to the observation that a salient feature of the product (here: aesthetic appeal) influences the perception of less salient features (here: perceived usability). This effect can be theoretically explained by attribution theory, which describes how individuals generalize from initial impressions to judgements about other properties (Fiske & Taylor, 1991; Weiner, 1985). Attribution theory predicts that humans make an internal attribution that the product was designed with care, competence, and quality, when it looks polished and aesthetically pleasing, extending that positive judgment to unrelated properties such as perceived ease of use and usability of the system. Hence, the aesthetic design does not merely please the eye, it triggers a broader causal inference that good-looking means well-made, so usability is perceived more favorably (Brun et al., 2021; Folkes, 1988; Weiner, 2000).

In contrast to the rather consistent pattern found for the influence of aesthetic appeal on perceived usability, the research literature does not provide an unequivocal picture for other components of the UX construct. Regarding user performance, three principal effect patterns of aesthetics were observed. (a) Users showed improved performance when an aesthetically appealing device was employed (e.g. Miller, 2011; Sonderegger & Sauer, 2010; Tuch et al., 2009). This effect sometimes only emerged under low usability (and high task difficulty) conditions (Moshagen et al., 2009; Reppa & McDougall, 2015). (b) User showed decreased performance when an aesthetically appealing device was employed (e.g., Ben-Bassat et al., 2006; Sauer & Sonderegger, 2009). (c) User performance was unaffected by the aesthetical appeal of the device (e.g., Chawda et al., 2005; Hartmann et al., 2007; Thüring & Mahlke, 2007). Most often, the positive link between aesthetics and behavioural measures of usability is explained by the interaction of affective, attentional, cognitive and motivational processes (for an overview, see Thielsch et al., 2019).

With regard to the effects of aesthetic appeal on affect, the positive affect mediation model (Norman, 2002) suggests positive consequences of aesthetics on emotions. However, the empirical evidence seems to be less clear, although several studies showed positive consequences of aesthetics on affective states of users (Reppa et al., 2021; Sonderegger et al., 2012; Thüring & Mahlke, 2007).

NON-VISUAL AESTHETICS IN UX EVALUATION

All these effects reported in the previous section referred to the influence of visual aesthetics, while the influence of non-visual aesthetics has been largely neglected. On the one hand, this is somehow understandable given the strong dominance of the visual sense of human perception (Schifferstein et al., 2010). On the other hand, there is also a need to obtain a better understanding of the role of the other human senses in the evaluation of UX.

In this respect, research has examined whether this robust halo effect of visual aesthetics would also be found for non-visual aesthetics (Sonderegger & Sauer, 2015). This work examined hearing and touch as two prominent non-visual senses. Surprisingly, no such influence was found in any of the three studies, in which three different consumer products were tested (smartphone, vacuum cleaner, video racing game). The results were explained by using the framework of sensory dominance (Fenko et al., 2010; Schifferstein, 2006).

In a study of Schifferstein (2006), the perceived importance of sensor modalities was examined for a set of different products analysing self-reports. For half of the products, vision came out as the most dominant sense, followed by touch, smell, hearing, and taste. Vision emerging as the predominant modality is explained by its ability to gather and process larger amounts of information about product relevant aspects in a shorter time, compared to other senses (Schifferstein & Cleiren, 2005). Despite this general visual dominance, the study by Schifferstein (2006) revealed for half of the products examined that one of the non-visual senses emerged as the predominant sense. These products had in common that the predominant sense was also the primary modality involved in product interaction (e.g., taste for foods and drinks), suggesting that the rank order is highly dependent on the type of product and its intended usage.

AESTHETICS OF TASTE

The sense of taste is an important function of control for ingestion, serving as cue in decision-making for food intake (Gibson, 1966). Furthermore, it is strongly

related to positive and negative emotions experienced when consuming food (Perullo, 2016). For the sensory modality of taste, a distinction between three principal elements can be made, which are taste qualities (bitter, sour, sweet, salty, umami), intensity of the stimulus, and its affective dimension (i.e. whether the taste is liked or not; Cardello & Wise, 2008). The gustatory sense cannot be considered independently from the olfactory sense as the latter influences the perceptual process of the former (Cardello & Wise, 2008). For example, the smell of a beverage influences its taste. For reasons of simplicity, we will refer to the concept of taste in this article acknowledging that olfactory sensory processes are involved in the sensory experience of tasting, too. In this context, we use the term 'aesthetics of taste' or 'appeal of taste' when referring to the experimental manipulation of taste whereas we use the term 'perceived gustatory attractiveness' when we refer to the effect of this manipulation on the (subjective) experience of a participant as a part of their UX.

Regarding the sensory dominance, taste is perceived as the most important sense for food products, followed by smell and vision (Schifferstein, 2006). The primary role of taste lies apparently in the interaction with food and is therefore generally of much smaller importance in user-product interaction than other sensory modalities (Schifferstein & Spence, 2008). However, the importance of taste may increase for the experience of systems that produce food or drinks (e.g., coffee machine). Some support for this idea may be found in the study of Fenko et al. (2010), in which participants often reported the dominant modality to be dependent on the product's primary function. For example, 'The most important is the smell and taste of the coffee because that is what a coffeemaker is for' (p. 37). This might indicate that regarding UX, it can be assumed that for the evaluation of a system producing food or beverages (e.g., coffee machine), taste is the dominant sensory modality. Therefore, referring to the halo effect and assumptions of the attribution theory described above, it can be expected that taste will show a similar effect pattern on UX measures when evaluating systems producing food or beverages as visual aesthetics for consumer goods where vision is the main sensory modality (e.g., mobile phones, webpages). This contrasts with sound and touch, which are rated most often second or third in importance for product evaluation (Schifferstein, 2006).

OUTPUT QUALITY AS A CONCEPT IN USER-DEVICE INTERACTION

As outlined in the previous section, it can be assumed that perceived quality of food or beverages is closely associated with the general experience of the user with the device. This includes consumer products used to produce food or drinks, which is immediately consumed after product utilisation (e.g., meal prepared in a multi-purpose kitchen appliance, espresso prepared by a coffee machine).

The importance of the outcome quality of a product produced by a technical device in UX evaluations has already been emphasised by Jordan (2020, see also; Hornbæk, 2006) when introducing the concept of 'quality of output'. In his framework of important usability variables, Jordan considers 'quality of output' to be an important effectiveness measure in product evaluation. When completing a task (e.g., preparing a good coffee), this measure indicates whether a user was successful or not in achieving the desired outcome (e.g., to drink a good coffee). Outcome quality can be assessed through subjective ratings by end users. A common approach is to collect users' self-reported satisfaction with the produced outputs, typically via rating scales administered after task completion (e.g., asking how pleasant the taste of the coffee is after taking a sip).

In principle, the concept of output quality should be independent of the concept of inherent usability (e.g., a coffee machine is easy to use but it still produces an appalling coffee). Despite the theoretical independence of the two concepts, it is not clear whether a user would make a clear distinction between output quality and usability features of the device. The spill-over of an effect to other UX measures may be described as a 'halo effect' (e.g., Fisicaro & Lance, 1990).

SYSTEM USABILITY

System usability is considered in this work as a product-inherent and objective characteristic that can be manipulated experimentally (as opposed to perceived usability which is the subjective consequence of the interaction of a person with an interface). The effects of good and poor system usability have been widely examined in research since it represents the essence of UX research. This generally shows degradations of user performance when low-usability devices are operated (e.g. induced by system delays; Barber & Lucas, 1983; Galletta et al., 2004; Sonderegger et al., 2019; Szameitat et al., 2009). These effects on user performance can also spread to non-performance variables such as user satisfaction (e.g., Moshagen et al., 2009) and emotion (e.g., Sonderegger et al., 2019; Thüring & Mahlke, 2007).

THE PRESENT STUDIES

The article presents three experiments[2] aimed to examine the influence of aesthetics of taste (conceptualised through 'quality of output') on typical measures of UX. Being the first studies that have investigated this issue, they aimed to determine whether aesthetics of taste is similar to visual aesthetics in its pattern of influence, following assumptions of the attribution theory, or shows a pattern more similar to non-visual aesthetics (e.g., sound and touch).

Based on the findings of the first study, we carried out a second one and, finally, a third study to determine the degree of consistency in the results pattern. In all three studies, the taste of coffee was manipulated when users operated a fully automatic coffee machine. Aesthetics of taste may represent a complex concept in the context of coffee making since it is influenced by a number of factors that influence aesthetics of taste and hence the gustatory experience of the coffee (Cameron et al., 2020). Several factors can be considered when manipulating the taste of a coffee. (a) The first group of factors refers to the ingredients for making coffee such as the quality of coffee beans and the quality of water. (b) The second group of factors refer to the setting of technical parameters of the coffee machine, such as pressure, temperature and grinding degree with which the coffee is processed, which influence the taste of the coffee served. (c) The third group of factors refers to physical features of the coffee, such as its temperature and strength (i.e. amount of grinded beans used to make the beverage). (d) The fourth group of factors refers to cognitive influences such as individual preferences (e.g., espresso is preferred to cappuccino), cognitive framing (e.g., a friend tells you this is an excellent coffee), and contrast effects (e.g., having had a bad coffee prior to this one). Based on these factors of influence, different approaches were employed in the studies presented in this article with the aim of contributing to the development of methods for experimentally manipulating aesthetics of taste. In all studies, a 2 × 2 experimental design was employed, with coffee taste being manipulated as the main independent factor and system usability as a secondary independent factor. In each study, we examined the influence of aesthetics of taste in a different way by varying the manipulation of taste.

System usability was included as a secondary independent factor as this allows us to determine whether the influence of aesthetics of taste would be different across different levels of task difficulty. This means we were primarily interested in an interaction between aesthetics of taste and system usability as previous research on visual aesthetics has shown that positive effects of aesthetics were mainly revealed under difficult conditions of use (Moshagen et al., 2009; Reppa & McDougall, 2015).

Based on the literature review, we assumed that a similar pattern of effects observed for visual aesthetics would also be found for aesthetics of taste for the UX evaluation of a coffee machine (see discussion on dominant sensory modality in section 1.3). Taste of the product is a very salient and important quality for the UX of a coffee machine, perhaps even more so than touch and sound for devices such as mobile phones, vacuum cleaners and steering wheels (cf. Sonderegger & Sauer, 2015). Therefore, we expected it to show a similar effect as visual aesthetics. More specifically, we predicted that high aesthetics of taste would have a positive effect on UX, compared to low aesthetics of taste. This hypothesis was put forward in all three studies. In each study, the same UX variables were measured, including perceived gustatory attractiveness, perceived usability, positive and negative affect, and performance.

EXPERIMENT I

GOALS OF STUDY

The first study can be considered a feasibility study using a smaller sample, addressing mainly main effects of independent variables. The aim of the experiment was to determine whether the manipulation of aesthetics of taste shows the expected effect in the manipulation check and whether it shows a WIBIG effect on UX (i.e. perceived usability, performance and affect) in much the same way as visual aesthetics was found to do. The research questions were examined by using an automatic coffeemaker.

METHOD

Participants

Due to the exploratory nature of the first study, a rather small sample of 40 participants was recruited without prior sample size estimation. Data collection was stopped after reaching the deliberately defined number of 40 participants. The sample consisted of psychology students (67.5% female), aged between 20 and 30 yrs (M = 22.95, SD = 2.46). Participants earned course credits in return for their participation. Participants rated their expertise of coffee below midscale (M = 8.20, SD = 4.51) on a scale ranging from 1 (neophyte) to 20 (connoisseur). When asked to rate how much they liked coffee, their rating was somewhat higher on the 20-point Likert scale (M = 12.83, SD = 4.86). More than half of the participants reported preferring milk in their coffee (N = 20) or to drink beverages like cappuccino (N = 6), latte macchiato (N = 3), whereas a minority reported preferring their coffee beverages without milk, like espresso (N = 4), black coffee (N = 3), ristretto (N = 1) or other ones (N = 3).

Experimental design

A 2 × 2 factorial between-subjects design was employed in this study. Aesthetics of taste (high vs. low) as the main independent factor was experimentally manipulated by offering coffee beans of low or high quality (see procedure and pilot study for details). System usability (high vs. low) as the subsidiary independent factor was experimentally manipulated by two means: (1) reducing readability and (2) reducing comprehensibility of the displayed content. Readability was reduced by setting the brightness of the display to the minimum level, which made it difficult to perceive the information provided on the display. Comprehensibility was reduced by replacing the text

labels of icons (i.e. coffee, espresso, tea) with numbers, leaving the participant with icons as the only means of identifying the displayed content. Thus, the manipulation consisted of offering a high-contrast display with labelled icons or a low-contrast display with unlabelled icons. A similar combination of means to increase task difficulty was already successfully implemented in a study of Sonderegger and colleagues (2012), in which the inherent system usability of mobile devices was manipulated.

Pilot study to test experimental manipulation of gustatory aesthetics

A pilot study was carried out to determine how aesthetics of taste could be best manipulated. It involved two tests (each $N = 5$) to obtain two coffees that most participants would consider as being of good and poor quality.

For the first test, two types of coffee were bought in a local store. The selection was based on price (cheapest vs. most expensive coffee). It was hypothesised that the cheapest coffee would be perceived as unappealing and the expensive one as appealing regarding taste. Five students were then recruited to taste and evaluate the taste of the two coffees in a within-subjects design with randomised order of presentation. A short break between the two coffees was scheduled which allowed the taste of the first coffee to be neutralised with mineral water. Using the one-item scale for assessing gustatory attractiveness (see measures and instruments), it emerged that the high-quality coffee ($M = 4.40$; $SD = 2.30$) was not rated much tastier than the low-quality coffee ($M = 3.00$; $SD = 2.83$; $t\ (4) = 1.30$; $p > .05$; $d = 0.61$) on a scale ranging from 1 (very bad) to 9 (very good). The post-experimental interview revealed that some participants had rated the high-quality coffee as bad as the low-quality one because it was too acid. Since the manipulation of the first test failed, we carried out a second test and modified our approach.

In the second test, we used the same coffees, but we added a small amount of liquid sugar to the appealing coffee (high-quality) to neutralize acidity and bitterness. Furthermore, we modified the grinding degree of the coffee machine for the unappealing coffee (low-quality) to make it more acid. Following the same procedure as in the first test, we tested again five students. Using the same scale for assessing gustatory attractiveness (see above), it emerged that the difference between appealing and the unappealing coffee was more pronounced, and the variability of the ratings was reduced ($M_{appealing} = 7.6$, $SD_{appealing} = 0.89$; $M_{unappealing} = 2.0$; $SD_{unappealing} = 0.71$). This effect did even reach significance level ($t\ (4) = 22.86$; $p < .001$; $d = 7.98$). We therefore considered this manipulation of aesthetics of taste to be suitable for the main study.

Measures and instruments

Gustatory attractiveness. Gustatory experience of the coffee was measured using a purpose-built one-item scale. The item referred to the gustatory attractiveness of the coffee ('Please rate your overall impression of the espresso as precise as possible!'). It was rated on a 9-point Likert scale (very bad – very good). The item was adapted from items used by Sonderegger and Sauer (2015) to measure non-visual attractiveness. It was also similar to an item that was used by the manufacturer of the coffee machine tested. The item was presented twice to the participants, at the beginning of the experiment (t_1; serving as a manipulation check) and at the end of it (t_2; examining the experimental effect). A single-item scale was chosen due to the repeated measurement of the construct. Previous research has shown that for overall ratings such as the one intended in this study, one-item scales show high correlations with multi-item scales assessing the same construct and involve less effort for the study participants and may reduce loss of motivation, dropouts, and inattentive responding (e.g., Fakunmoju, 2020; Giraldo-O'Meara et al., 2014; Nagy, 2002; see also Allen et al., 2022).

Perceived usability. The System Usability Scale (SUS; Brooke, 1996) was used to measure perceived usability. Using a 5-point Likert scale, it consists of 10 items and represents a well-established instrument for assessing system usability. A meta-analysis based on aggregated data revealed a high internal consistency of the questionnaire (Cronbach's $\alpha > .91$, Bangor et al., 2008). Internal consistency of the used scale was found to be good (study I: $\alpha = .82$; study II: $\alpha = .89$; study III: $\alpha = .88$).

Affect. User affect was assessed using two instruments: Self-Assessment Manikin (SAM; Bradley & Lang, 1994) and the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF; Thompson, 2007). To obtain a baseline measure of affect, each questionnaire was administered before interacting with the system (t_1) and again after test completion (t_2).

The SAM measures three dimensions of affect (pleasure-displeasure, degree of arousal, and dominance-submissiveness) that are based on the semantic differential scale by Mehrabian and Russel (1974) and are considered independent of each other. These dimensions are graphically represented in the form of manikins and each of them is rated on a nine-point scale. Although the SAM includes the three dimensions valence, arousal, and dominance, only valence and arousal were assessed in the present study. Previous research has shown that dominance often contributes less to the explanation of affective responses and tends to exhibit weaker and less consistent effects than valence and arousal. Furthermore, the dominance construct is often more difficult for participants to interpret, which can reduce measurement reliability.

In addition, affect was measured using the I-PANAS-SF (Thompson, 2007), which is a short version of the Positive and Negative Affect Schedule (PANAS), originally developed by Watson et al. (1988). It consists of 10

adjectives measuring positive affect (PA) such as 'being active' and negative affect (NA) such as 'being upset'. Responses are made on a 5-point Likert scale (not at all – extremely). For the purpose of this experiment, the wording of the validated German version was used (Krohne et al., 1996). They reported that internal consistency of their German scales was found to be good, with a Cronbach's alpha of .84 for both dimensions. Internal consistency of the used scale was found to be acceptable for the positive dimension (study I: α = .80; study II: α = .75; study III: α = .86), but was somewhat lower for the negative dimension (study I: α = .62; study II: α = .62; study III: α = .80). Taking these low values of internal consistency into consideration, results of NA should be interpreted with caution.

Performance. Two efficiency measures were used. Task completion time (s) was recorded by the experimenter using a stopwatch, and the number of user interactions was measured by counting each user manipulation of the system such as the use of manual controls. The counting was done by the experimenter by using a tally, with the entire interaction being recorded by a camera so that these measurements could be verified.

Materials

A fully automatic coffee machine (Jura IMPRESSA Z9) was used. It was equipped with a complex menu structure in the form of a graphical interface, which allowed the experimenter to develop sufficiently complex task scenarios. A camera was employed to make video and audio recordings of the interaction (Logitech Webcam C310) to be able to verify the assessment of efficiency indicators.

User tasks

Participants were asked to perform seven tasks during the experiment. All tasks were used for the efficiency measurement except the final task. Its purpose was to allow participants to make their final rating of coffee quality.

The following seven tasks were set: (1) Change the default screen wallpaper to one of your choice; (2) Find out how many portions of milk have already been prepared by the machine; (3) Change screen wallpaper to default settings; (4) Change settings to energy saving mode; (5) Find out whether some maintenance work is required on the machine; (6) Change settings from energy saving mode to default settings; and (7) prepare an espresso using the main menu. For the analysis of efficiency, data from tasks 1–6 were aggregated.

Procedure

The experiment was carried out in a UX laboratory. Participants were randomly assigned to one of the experimental conditions. After having welcomed the participant, the experimenter explained that the purpose of the study was to conduct a UX evaluation of a coffee machine, followed by typical instructions used in user evaluation studies (e.g., 'we do not test your abilities but the product', 'please try to complete the tasks as well as possible without help'). Participants were asked for their consent to videotape the operation of the coffee machine using a camera recording only hand movements on the interface of the coffee machine. Following these instructions, an informed consent form was completed and signed. Participants were then asked to complete the following questionnaires: a demographic questionnaire, a questionnaire about consuming habits, the PANAS and the SAM. This was followed by offering them a first coffee of their choice (e.g., americano, espresso, ristretto), which they were asked to taste and rate afterwards using the gustatory attractiveness scale. Participants were then given the seven tasks (see above), which they were asked to carry out on the automatic coffee machine. The last task involved making an espresso, which they were asked to taste and rate afterwards using the gustatory attractiveness scale. Please note that every participant in the same experimental condition tasted a coffee of the same quality. As we expected that the (manipulated) taste would influence participant behaviour (WIBIG hypothesis), the manipulation needed to be applied before participants interacted with the coffee machine. To evaluate whether the interaction with the coffee machine had an influence on the evaluation of gustatory attractiveness, it was important to measure this variable a second time after the interaction of participants with the coffee machine. After tasting the second coffee, participants completed several questionnaires (i.e., PANAS, SAM and SUS). Finally, participants were given the opportunity to provide feedback to the experimenter in the post-experimental interview, followed by a debriefing. Participation in the study took about 45 minutes.

Data analysis

Due to the small sample size, only main effects were calculated using Student/Welch t-tests. For measures of affect, a one-factorial analysis of covariance was carried out, with the baseline measure being entered as a covariate. JASP software (JASP Team, 2022) was used for data analysis. The estimation and interpretation of effect sizes was based on Lenhard and Lenhard (2016). The statistical analysis plans were not pre-registered but followed strictly the hypotheses that were formulated prior to data collection.

RESULTS

Manipulation checks

Aesthetics of taste. The manipulation check for pre-test coffee rating (t_1) revealed that participants rated the gustatory attractiveness of the appealing coffee as being tastier (M = 4.95, SD = 1.96) than the unappealing coffee (M = 3.85, SD = 2.06). Statistical analysis showed

that the significant (one-tailed) effect was of a medium size (t (38) = 1.731; p = .046, d = 0.547 [1.077 – 0.017]). A smaller (and not significant) effect was observed for the second coffee rating (t_2), with higher ratings for the appealing coffee (M = 4.55, SD = 1.67) than for the less appealing coffee (M = 3.75, SD = 2.10) (t (38) = 1.33; p = .095, d = 0.422 [0.948 – –0.104]).

System usability. The manipulation check revealed higher usability ratings for the high usability condition compared to the low usability for the SUS scale, but the small effect did not reach significance level (M_{high} = 67.13; SD = 10.71 vs. M_{low} = 63.50; SD = 20.11; t (38) = 0.71, p = .24, d = 0.23 [0.747 – –0.297]).

Evaluation of main effects for WIBIG hypothesis

To test the WIBIG-hypothesis, the influence of aesthetics of taste on the different UX measures was analysed. The descriptive data for the different dependent variables as well as the t-statistics are presented in Table 1. Overall, the data analysis revealed that all effects ranged from very small to small and none reached significance.

The influence of usability on perceived gustatory attractiveness revealed a small non-significant effect (M_{high} = 4.60, SD = 1.88 vs. M_{low} = 4.20; SD = 2.26; t (38) = 0.61, p = .28, d = 0.19 [0.714 – –0.329]).

DISCUSSION

The most important finding was that the manipulations of aesthetics of taste did not show the expected large effects in the manipulation check with only small to medium effect sizes being found. This indicates the difficulty of implementing manipulations of aesthetics that work consistently. Similar problems were already observed in previous work that aimed to manipulate visual aesthetics (Sonderegger, Uebelbacher, et al., 2014).

The difficulty of achieving a successful manipulation of aesthetics is an interesting finding. It indicates that across many sensory modalities (if not all), there are considerable interpersonal differences, which may make it difficult to find a manipulation of aesthetics that is generally perceived as corresponding to high or low appeal. The difficulties reported here and in other work are somewhat at odds with the claims of the existence of general rules of aesthetics (e.g., Hekkert & Leder, 2008). However, it may be that beyond these general rules, many other factors influenced the perception of aesthetics, but these factors are related to individual preferences. Given the difficulties of achieving a successful manipulation of aesthetics of taste, we wished to replicate the study by slightly modifying the way aesthetics of taste was manipulated.

Regarding the manipulation of system usability, the manipulation check indicated only a small effect on the SUS scale. Comparable effects have been reported in a similar manipulation of system usability implemented in a previous study (Sauer & Sonderegger, 2022).

In addition, data analysis did not reveal an effect of aesthetics of taste on any of the main UX measures (i.e. perceived usability, efficiency and affect). The small sample size (post-hoc analyses revealed a statistical power of .29 (α = .05) for perceived usability), indicates a very high risk of Type II errors. Nevertheless, the overall pattern of results can be interpreted as a first indicator that the effects of gustatory aesthetics might be more similar to other types of non-visual aesthetics such as sound and touch (Sonderegger & Sauer, 2015) than to visual aesthetics. To test these hypotheses under more robust conditions, a second study was conducted in which the sample size was increased, and the experimental manipulations of aesthetics of taste and system usability were adjusted to increase the effect of the experimental manipulation.

EXPERIMENT II

GOALS OF STUDY

The second experiment also aimed to determine whether aesthetics of taste influences measures of UX that are conceptually independent of aesthetics. Due to small effect sizes obtained in the first experiment, we aimed to increase the power of the test by using a larger sample size. We also used a different approach to manipulate aesthetics of taste given the small effects

MEASURES	HIGH AESTHETICS OF TASTE *M (SD)*	LOW AESTHETICS OF TASTE *M (SD)*	INFERENTIAL STATISTICS *t/F p d/η^2*
Perceived usability (0–100)	68.13 (15.19)	62.50 (16.68)	1.12 .14 0.35 [1)]
Positive affect (5–25)	17.05 (2.98)	16.40 (3.36)	0.50 .48 .006 [2)]
Negative affect (5–25)	6.60 (1.96)	6.55 (2.44)	0.07 .79 .000 [2)]
Valence (1–9)	6.40 (0.82)	6.85 (0.93)	1.39 .25 .030 [2)]
Arousal (1–9)	4.15 (1.81)	4.35 (1.69)	0.00 .98 .000 [2)]
Task completion time (s)	304.80 (92.28)	314.65 (142.00)	0.26 .60 0.08 [1)]
User interactions (number)	82.75 (14.25)	89.45 (21.58)	1.16 .13 0.37 [1)]

Table 1 UX measures as a function of aesthetics of taste (M = mean; SD = standard deviation).

[1)] t (38) p d; [2)] F (1, 37) p η^2 with the measure of t1 (baseline) used as covariate.

in the manipulation checks in the previous study. This was done by manipulating taste based on individual user preferences that were assessed prior to the operation of the coffeemaker. As in the preceding study, we also aimed to examine whether the effects of aesthetics of taste would hold under different levels of system usability. Since the usability manipulation did not show strong effects in the first study, we also adapted the manipulation of this independent variable.

METHOD

Participants

Sixty students (81.7% female) reading psychology were recruited for this experiment, with their ages ranging from 18 to 29 years (M = 20.83, SD = 1.70). In return for their participation, they earned course credits. For the a priori sample size estimations, we referred to previous work to define an expected effect since no comprehensive meta-analysis has been published so far in this domain (Anderson, 2019). As we expected similar effects as for visual aesthetics (i.e. f = .37; see discussion section in Sonderegger & Sauer, 2015 for more details on effect sizes of studies addressing the influence of visual aesthetics on UX measures), our sample size estimation resulted in a required sample of 60 participants (with f = .37, α err prob = .05, power = .80; for test-economic reasons, a beta error rate of 20% seemed acceptable, especially since additional studies addressing this topic were being planned).

Participants rated their expertise of coffee below midscale (M = 7.82, SD = 4.56) on a scale ranging from 1 (neophyte) to 20 (connoisseur). When asked to rate how much they liked coffee, their rating was somewhat higher on the 20-point Likert scale (M = 12.85, SD = 5.06). Most of the participants reported preferring to add milk to their coffee (21) or to drink similar beverages like cappuccino (14), latte macchiato (11), whereas a minority reported preferring their coffee beverages without milk, like espresso (5) or black coffee (9).

Regarding their expertise in using coffee machines they gave a rating below midscale (M = 6.60; SD = 3.72) on the 20-point Likert scale. Forty-eight participants were in possession of a coffee machine of some kind: 10 used a fully automatic coffee machine, 15 used a capsule coffee machine, 14 used a moka pot, 2 used a filter coffee machine and 7 used a French press.

Experimental design

As in the previous experiment, a 2 × 2 factorial between-subjects design was employed. Aesthetics of taste as the main independent factor was experimentally manipulated by varying the extent to which participants like the coffee being served (favourite coffee vs. least preferred coffee, see method section of Experiment I for a more detailed description of the procedure to manipulate aesthetics of taste). System usability (high vs. low) as the subsidiary independent factor was experimentally manipulated by adaptation of system reliability. A combination of unexpected system shutdowns and an impossible task was used in the low-usability condition (see method section of Experiment I for more details). This revised approach to manipulating system usability was implemented because the manipulation employed in the previous study did not show the anticipated effects. Therefore, a more pronounced usability manipulation was introduced to achieve effect sizes comparable to those observed in the taste manipulation and in prior studies examining visual aesthetics.

Measures and instruments

The following UX components were assessed: Perceived gustatory attractiveness, perceived usability, affect and efficiency. This corresponds to the same list of dependent variables measured in experiment I, where a description of the instruments and measurement procedures may be found.

Materials

The same coffeemaker was used as in the previous study (i.e. Jura IMPRESSA Z9). We also used the same camera to obtain recordings of user-product interaction (i.e. Logitech Webcam C310).

User tasks

Participants were asked to perform nine tasks using the coffee machine. Four of the tasks were used for the efficiency measurement (tasks 3, 4, 6, 7). The others served either as a warm-up task (task 1), as part of the experimental manipulation of system usability (i.e. tasks 2, 5, 8), or to allow participants to make their final rating of coffee quality (task 9).

The following nine tasks were set: (1) Fill water into the machine up to the required level (warm-up task); (2) Change the default screen wallpaper to one of your choice (power failure in low-usability condition); (3) Check whether some maintenance of the machine is required; (4) Change screen wallpaper to default settings; (5) Change settings to energy saving mode (power failure in low-usability condition); (6) Find out how many portions of milk have already been prepared by the machine; (7) Disable settings of energy saving mode; (8) Check in the menu whether it is necessary to add more coffee beans (impossible task was only presented in low-usability condition); and (9) prepare a specific coffee beverage using the main menu (most or least preferred beverage). For the analysis of efficiency, data of tasks 3, 4, 6 and 7 were aggregated.

Procedure

In a first step, participants' preferences were assessed regarding different coffee beverages. Seventy-one students took part in this assessment. They completed

an online questionnaire where they had to make a ranking of five different coffee beverages (black coffee, coffee with milk, latte macchiato, cappuccino, espresso), ranging from the most preferred to the least preferred one. This ranking was made separately for the morning and afternoon coffee. In addition, participants gave a rating of how much they like drinking each of the five beverages (on a 1–50 Likert scale) and whether they prefer drinking it with or without sugar. The preferences of the participants were characterised by considerable interpersonal differences. These data were then used to determine for each participant their most preferred and least preferred coffee in the morning and in the afternoon. Eleven participants were excluded because the difference between most preferred and least preferred beverage was not very pronounced, resulting in sixty participants for the main test.

The experimental procedure for the main test was very similar to the one used in the previous study (see procedure of study I). The only difference regarding the manipulation of gustatory aesthetics was that participants had to prepare their preferred (or non-preferred) coffee for the gustatory rating. They were allowed to drink their favourite coffee in the high-gustatory aesthetics condition (e.g., with sugar), whereas they were asked to drink their least preferred beverage (e.g., without sugar if they did like sugar) in the low-gustatory aesthetics condition.

Compared to the procedure in study I, the usability manipulation was implemented differently, since it only showed small effects on the manipulation check in the preceding study. In the low-usability condition, a power failure was initiated by the experimenter (via a remote control hidden in his hand), which led to a system shutdown of the coffee machine. This created the impression among participants that a software failure (when clicking on a certain command in the menu) caused this system shutdown. In addition, participants were given a task that could not be solved, which was also expected to decrease usability evaluations of the device.

Data analysis

A 2-factorial model of analysis of variance was used to test the hypotheses. Analysis of covariance was carried out for the outcome variables of affect, with the baseline measure being entered as a covariate. Student/Welch t-tests were used for manipulation checks. In order to test for equivalence of non-significant effects (i.e., H1 assumes that gustatory aesthetics has not the same effect on UX as visual aesthetics), the *two one-sided t-tests* (TOST) procedure for equivalence tests (Lakens, 2017; Lakens et al., 2018) was applied, with a smallest effect size of interest (SESOI) defined based on previous research on visual aesthetics (Welch's t test for independent samples with equivalence bounds of $\Delta L = -0.74$ and $\Delta U = 0.74$, see discussion section in Sonderegger & Sauer, 2015 for more details on effect sizes of studies addressing the influence of visual aesthetics on UX measures). G*Power (Faul et al., 2007) was used for sample size estimation. The estimation and interpretation of effect sizes was based on Lenhard and Lenhard (2016). The statistical analysis plans were not pre-registered but followed strictly the hypotheses that were formulated prior to data collection.

RESULTS

In order to keep the presentation of results clear and concise, graphs that visually represent the statistical comparisons (raincloud, box and violin plots) are included in the supplemental material (Figures 2 to Figures 11 in the supplementary material section).

Manipulation checks

Aesthetics of taste (cf. Figure 2 in the supplementary material section). The successful manipulation check for the pre-test coffee rating (t_1) revealed that participants rated the appealing coffee as being tastier (M = 6.47, SD = 1.74) than the unappealing coffee (M = 5.33, SD = 2.04). The statistical analysis showed that the significant (one-tailed) effect was of a medium size ($t(56.6) = 2.32$; $p = .012$, $d = 0.60$). A larger effect was observed for the second coffee rating (t_2), with higher gustatory attractiveness ratings for the appealing coffee (M = 6.70, SD = 1.56) than for the less appealing coffee (M = 4.90, SD = 2.19) ($t(52.39) = 3.67$; $p < .001$, $d = 0.95$).

System reliability (cf. Figure 3 in the supplementary material section). The successful manipulation check revealed a medium to large effect, with higher perceived usability ratings (SUS scale) for the high reliability condition compared to the low reliability (M_{high} = 64.08; SD = 19.07 vs. M_{low} = 50.33; SD = 19.52; t (58) = 2.76, p = .004, d = 0.71).

Analysis of data testing the WIBIG hypothesis

The WIBIG hypothesis suggests that aesthetics of taste has a similar positive influence on UX ratings as this has been shown for visual aesthetics. In addition, it was expected that this effect is especially pronounced for interactions with systems of low usability for measures of efficiency. Data analyses for the individual dependent variables revealed the following results.

Perceived usability. While the effect of system reliability reached significance level ($F = 7.62$; $p < .01$; $\eta^2_{partial} = .12$), only very small nonsignificant effects of aesthetics of taste on perceived usability ($F < 1$; $\eta^2_{partial} = .02$, *cf. Figure 4 in the supplementary material section*) and of the interaction ($F_{\text{aesthetics of taste} \times \text{system usability}} = 1.18$; $df = 1, 56$; $p = .28$; $\eta^2_{partial} = .02$) could be observed, indicating that the WIBIG hypothesis was not confirmed for measures of perceived usability. The results of the equivalence tests indicated that the non-significant effect is statistically equivalent ($t(56.56) = 1.995$, $p = .025$).

Affect. Controlling for the baseline measure, data analysis revealed no significant effect of aesthetics of taste on positive affect ($M_{appealing}$ = 17.20; *SD* = 2.17; $M_{unappealing}$ = 16.40; *SD* = 3.32; *F* (1, 55) = 3.54; *p* = .065; $\eta^2_{partial}$ = .06). For negative affect, gustatory aesthetics also showed a very small nonsignificant effect (*F* (1, 55) = 2.62; *p* = .11; $\eta^2_{partial}$ = .045; *cf. Figure 5 in the supplementary material section*). The results of the equivalence tests on a change score (affect at t2 minus baseline measure of affect) indicated that the effect is inconclusive for positive affect (*t*(57.79) = –1.613, *p* = .112) but statistically equivalent for negative affect (*t*(57.41) = –2.515, *p* = .007). In addition, system reliability showed no effect on both dimensions of affect (F_{pos} < 1, $\eta^2_{partial}$ = .01; F_{neg} < 1, $\eta^2_{partial}$ = .01). Also, the interaction between the two independent variables did not reveal any effect (F_{PA} < 1, $\eta^2_{partial}$ = .01; F_{NA} < 1, $\eta^2_{partial}$ = .01).

Valence and arousal. Analysis of covariance revealed a small significant main effect of aesthetic appeal on valence ($F_{valence}$ (1, 55) = 5.02, p = .029, $\eta^2_{partial}$ = .08, see Figure 1), with ratings of participants in the gustatory pleasing condition being higher (*M* = 6.77; *SD* = 1.14) than in the displeasing condition (*M* = 6.47; *SD* = 1.89). Aesthetic appeal did not show an effect on arousal ($F_{arousal}$ (1, 55) = 1.87, p = .18, $\eta^2_{partial}$ = .03). The results of the equivalence tests on a change score (arousal at t2 minus baseline measure of arousal) indicated that the effect is statistically equivalent (*t*(57.99) = –1.863, *p* = .034). There was no effect of system reliability ($F_{valence}$ < 1, $\eta^2_{partial}$ = .01; $F_{arousal}$ < 1, $\eta^2_{partial}$ = .00). There was no significant interaction between aesthetic appeal and system reliability, neither for valence, nor for arousal ($F_{valence}$ < 1, $\eta^2_{partial}$ = .00; $F_{arousal}$ < 1, *df* = 1, 36, *p* > .05, $\eta^2_{partial}$ = .01).

Task completion time. Data analysis for the average time needed to complete a task (without considering the system shutdowns and the impossible task in the low system reliability condition) revealed no differences between experimental conditions ($F_{aesthetics\ of\ taste}$ < 1; $\eta^2_{partial}$ = .01; $F_{system\ reliability}$ (1, 56) = 1.19; *p* = .28; $\eta^2_{partial}$ = .02) and showed that the interaction was not significant either (*F* < 1; $\eta^2_{partial}$ = .00; *cf. Figure 6 in the supplementary material section*). The results of the equivalence tests indicated that the effect for aesthetics is statistically equivalent (*t*(56.51) = –2.308, *p* = .012).

Number of user inputs. Data analysis revealed that aesthetics of taste had no effect (*F* < 1; $\eta^2_{partial}$ = .00; *cf. Figure 6 in the supplementary material section*), on number of user inputs, with results of the equivalence tests indicating that the effect for aesthetics is statistically equivalent (*t*(53.67) = –2.618, *p* = .006).

System reliability as manipulation of usability showed a medium effect on number of user inputs (*F* = 5.21; *df* = 1, 56; *p* = .03; $\eta^2_{partial}$ = .09). Participants in the low-reliability condition needed significantly more inputs to complete the tasks (*M* = 60.33, *SD* = 16.84) than participants in the high-reliability condition (*M* = 51.53, *SD* = 12.84). No significant interaction between both independent factors was found (*F* = 2.26; *df* = 1, 56; *p* = .14; $\eta^2_{partial}$ = .04).

DISCUSSION

The main finding of Experiment II was that the WIBIG hypothesis suggesting an effect of aesthetics of taste on UX measures had not been confirmed, despite the successful experimental manipulation of aesthetics of taste showing a medium (t_1) to large (t_2) effect on the measure of perceived gustatory attractiveness. This suggests that we can state with more confidence that the nil effects observed for UX measures are indicative of an insignificant WIBIG effect of gustatory aesthetics (as compared to visual aesthetics), which is also confirmed by most of the equivalence tests. The unique small but significant effect found for valence might be an incidental result, as the effect was not confirmed by the PA and NA measures and hence needs confirmation in additional studies.

The adapted manipulation of usability, which consisted in the adaptation of the reliability of the coffee machine with an unexpected shutdown and the implementation of an impossible task, showed a larger effect regarding the evaluation of perceived usability compared to experiment I. The effect of the manipulation check still was smaller compared to the effect of the manipulation of aesthetics of taste. Several options to increase the effect of the usability manipulation were discussed and pre-tested by the authors. After careful consideration, we decided to use the same manipulation as in Experiment II. The primary reason for this decision was that any modifications to the display, such as changes in contrast or the use of different visual materials, would also have carried the risk of influencing the visual aesthetics of the appliance. This, in turn, could have introduced a problematic confound, making it difficult to disentangle the effects of the manipulation from those of visual appeal.

However, the effect sizes of manipulations of visual aesthetics in previous research showed larger effects (Sonderegger & Sauer, 2015) as we managed to obtain for aesthetics of taste in this study. Therefore, it needs to be tested whether such a result can be corroborated by increasing the strength of the manipulation of gustatory aesthetics.

EXPERIMENT III

GOALS OF STUDY

As a consequence of the results of the two first experiments regarding the WIBIG hypothesis, we changed the set up of the study again to examine whether the nil effect of gustatory aesthetics would also emerge in a third study by increasing the effect of the gustatory

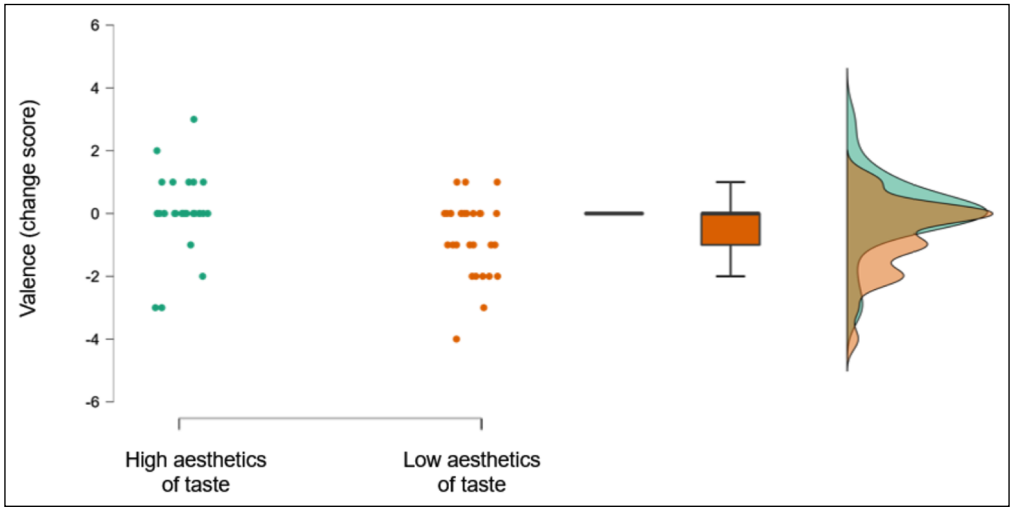


Figure 1 Raincloud, box and violin plots of valence ratings (change score of t2 minus baseline measure) as a function of aesthetics of taste.

aesthetics manipulation. The third study was conducted to determine whether the use of a more homogeneous user group of self-selected coffee experts would provide similar results. The use of a homogeneous group was expected to increase the power of the experimental manipulation. The use of self-selected coffee experts would also allow us to test a group of users who were expected to be more knowledgeable about coffee than the users in the two previous experiments. The manipulation of aesthetics of taste was modified by making use of the contrast effect (i.e. contrasting the quality of the coffee by offering a 'contrast coffee' prior to the coffee that was to be assessed; i.e. the 'contrast coffee' was very good in the low-gustatory aesthetics condition, and very poor in the high-gustatory aesthetics condition). The contrast effect is also known in cognitive psychology, where it refers to a contextual bias effect describing the influences on the assessment of a target after a background target had been presented (Chen & He, 2016).

Again, as in the second study, we aimed to examine whether the effects of aesthetics of taste would hold under different levels of system usability. We used the same usability manipulation as in the preceding study where it had worked well.

METHOD

Participants

Thirty-eight students (73.3% female) were recruited for this experiment, with their ages ranging from 19 to 37 years (M = 23.40, SD = 3.39). They earned 10 Swiss Francs in return for their participation. One participant had to be excluded from the analysis of efficiency measures because of a corrupted video recording. For the a priori sample size estimation we referred to experiment II. This sample could not be attained as a consequence of the difficulty in recruiting coffee experts as well as some unexpected constraints regarding the availability of the coffee machine. Data collection had to be stopped because we had to return the coffee machine. We are aware of this limitation and therefore interpret the results of this study with some caution.

Participants were self-selected coffee experts who considered themselves knowledgeable about what constitutes a good espresso. They also reported drinking espresso regularly. They were recruited via flyers and by an email that was sent out to all students of the host University.

Participants rated their expertise of coffee above midscale (M = 11.47, SD = 4.30) on a scale ranging from 1 (neophyte) to 20 (connoisseur). When asked to rate how much they liked coffee, their rating was high on the 20-point Likert scale (M = 17.66, SD = 2.57).

With regard to their expertise in using coffee machines, they provided a rating above midscale (M = 11.16; SD = 3.89) on the 20-point Likert scale. Thirty-six participants were in possession of a coffee machine: 6 participants used fully automatic coffee machine, 15 a capsule coffee machine, 12 a moka pot and 3 used a French press.

Experimental design

A 2 × 2 between-subjects design was employed, with aesthetics of taste (high vs. low) and system usability (high vs. low) being manipulated as independent variables. Aesthetics of taste was manipulated by using freshly roasted beans for the aesthetically pleasing coffee, and two-year old beans from the same brand for the unappealing coffee. In addition, participants were confronted with both coffees and were asked to rate their coffee in comparison to the contrast coffee (for more details see pilot studies). System usability was manipulated in the same way as in experiment II, that is, by adapting system reliability with inducing shutdowns during system use and by giving an impossible task.

Pilot studies to test experimental manipulation of gustatory aesthetics

In the two previous studies, we used two very different approaches to manipulate aesthetics of taste: group-

based approach vs. person-based approach. In the group-based approach, we aimed to create two coffees that were considered by a large group of people either of good or bad taste (i.e. corresponding to the 'it is a matter of taste' idiom). In the person-based approach, we aimed to identify the personal preferences of participants and offered them a coffee that they would consider either as good or bad. The results of the manipulation check in study I showed that a strong effect for the manipulation of aesthetics of taste was rather difficult to obtain when using a group-based approach. Therefore, we examined further possibilities to manipulate aesthetics of taste using the group-based approach. In a first step, we added grapefruit concentrate to make the unappealing coffee used in study I worse. Although this manipulation was successful, we did not pursue this idea because the coffee was perceived as too artificial in taste.

We then approached experts from a Swiss coffee roasting company who provided us with a high- and low-quality coffee. In a first pilot study (N = 13), participants rated both coffees (order was counterbalanced) in a within-subjects design, with a short break for neutralizing the taste with water. The analysis of all ratings (i.e. first and second coffee) revealed that participants reported the taste of the appealing coffee to be significantly better (M_{A+} = 5.08, SD = 2.33) than of the unappealing coffee (M_{A-} = 2.69, SD = 1.32; $t(12)$ = 4.34; $p < .001$; d_{RM} = 1.06). However, we were concerned that similar problems could occur as in study I, where the large effects of a successful pilot test (adopting a repeated-measures design) could not be replicated in the main study (adopting a between-subjects design). Therefore, we analysed only the first rating (as it would be the case in the main study). The results suggested that participants rated both coffees very similarly (M_{A+} = 3.67, SD = 2.34; M_{A-} = 3.29, SD = 1.11; $t(11) = 0.385$; $p = .71$; $d = 0.21$), indicating that without the presentation of a contrast coffee, the manipulation of taste was not successful. Therefore, we used a different manipulation of gustatory aesthetics to further increase the strength of the manipulation.

With a second pilot study (N = 12), adopting the same experimental design, we compared freshly roasted coffee with coffee made of beans that were roasted 4 years ago (i.e. taste and aroma was lost). This second pilot study revealed very similar results than the one reported above: The analysis of all ratings (i.e. first and second coffee) revealed that participants reported the taste of the appealing coffee to be significantly better (M_{A+}=5.75, SD=1.60) than the unappealing coffee (M_{A-}=2.92, SD=1.83; $t(11) = 3.36$; $p = .006$; $d_{RM} = 1.67$). When analysing only the first coffee served, the results suggested that participants did not perceive the two coffees to be considerably different in taste (M_{A+} = 5.50, SD = 1.76; M_{A-} = 4.00, SD = 1.79; t (10) = 1.464; p = .17; d = 0.85). However, a large (significant) difference in perceived gustatory attractiveness was achieved when analysing the second coffee served (M_{A+} = 6.00, SD = 1.55; M_{A-} = 1.83, SD = 1.17; t (10) = 5.259; $p < .001$; d = 3.04) as a between-subjects variable. Based on these findings we decided to include the use of a contrast coffee in experiment III to maximise the effect of the experimental manipulation of aesthetics of taste.

Measures and instruments

The following UX components were assessed: Perceived gustatory attractiveness, perceived usability, affect and efficiency. This corresponds to the measures used in the two other experiments. A description of the instruments and measurement procedures may be found in the method section of experiment I.

Materials

In this experiment, the same coffee machine was used as in the two previous studies (i.e. Jura IMPRESSA Z9). Again, we used the same camera to obtain recordings of user-product interaction (i.e. Logitech Webcam C310).

User tasks

Participants were asked to perform eight tasks using the coffee machine. Again, four of the tasks were used for efficiency measurement (tasks 2, 3, 5, 6). The others served either as part of the experimental manipulation of system usability (i.e. tasks 1, 4, 7), or to allow participants to make their final rating of coffee quality (task 8).

The following eight tasks were set: (1) Change the default screen wallpaper to one 'blue sky' (power failure in low-usability condition); (2) Check in the information system whether some cleaning of the coffee machine is required; (3) Change screen wallpaper to default settings; (4) Change the default shutdown of the machine from a 120-min interval to a 30-min interval (power failure in low-usability condition); (5) Find out how many portions of espresso have already been prepared by the machine; (6) Change the default shutdown of the machine back from a 120-min interval to a 30-min interval; (7) Check in the menu whether it is necessary to add more coffee beans (impossible task was only presented in low-usability condition); and (8) prepare an espresso using the main menu. For the analysis of efficiency, data of tasks 2, 3, 5 and 6 were aggregated.

Procedure

The experimental procedure was very similar to the one used in experiment I. The main difference consisted in the manipulation of aesthetics of taste, for which a 'contrast coffee' was offered in addition to the good or bad coffee at t1. This means that in the beginning of the experiment, two coffees were presented (prepared with two machines of the same make) to the test participants. The first coffee representing the contrast stimulus was very different in taste from the second

coffee representing the experimental manipulation (i.e. if the experimental manipulation required a tasty coffee, the contrast coffee was unappealing, and vice versa). At the end of the experiment, a third coffee had to be rated. It was of the same taste as the 'second coffee' served at the beginning of the experiment (e.g., both were gustatory appealing). Again, as in the first experiment, every participant in the same experimental condition tasted and rated the two coffees of the same quality (either gustatory appealing or unappealing). Regarding the manipulation of system usability, we used the same approach as in experiment II.

Data analysis

Due to the small sample size, only main effects were calculated using Student/Welch t-tests. For measures of affect, a one-factorial analysis of covariance was carried out, with the baseline measure being entered as a covariate. The estimation and interpretation of effect sizes was based on Lenhard and Lenhard (2016). The statistical analysis plans were not pre-registered but followed strictly the hypotheses that were formulated prior to data collection.

RESULTS

Manipulation checks

Aesthetics of taste. For the pre-test coffee rating (t_1), the manipulation check was successful: participants rated the appealing coffee as being tastier ($M = 6.16$, $SD = 1.39$) than the unappealing coffee ($M = 3.90$, $SD = 2.45$). Statistical analysis showed a large significant (one-tailed) effect ($t(28.46) = 3.51$; $p < .001$, $d = 1.14$). For the second coffee rating (t_2), the experimental manipulation of aesthetics of taste led to higher attractiveness ratings for the appealing coffee ($M = 5.95$, $SD = 2.01$) than for the less appealing coffee ($M = 3.79$, $SD = 2.35$), which resulted in a large effect (t (36) = 3.04; $p = .002$, $d = 0.99$).

System reliability. The successful manipulation check revealed a large effect, with higher usability ratings for the high reliability condition compared to low reliability for the SUS scale ($M_{high} = 79.34$, $SD = 13.46$ vs. $M_{low} = 59.47$, $SD = 17.17$; t (36) = 3.97, p < .001, $d = 1.29$). These results indicate that the manipulation of aesthetics of taste and system usability were leading both to large effects and were successful.

Analysis of data testing the WIBIG hypothesis

Perceived usability. The analysis showed that there was no significant effect of aesthetics of taste (t (36) = 0.29, p = .39, $d = 0.09$), indicating that the WIBIG hypothesis was not confirmed for measures of perceived usability (*cf. Figure 7 in the supplementary material section*).

Affect. Analysis of covariance showed that the manipulation of gustatory aesthetics had no effect on both dimensions of affect (F_{PA} (1, 35) = 1.09, $p = .30$, $\eta^2_{partial} = .03$; F_{NA} (1, 35) < 1, $\eta^2_{partial} = .01$). Exclusion of one extreme value for NA in the high gustatory aesthetics group did not change considerably the result of the ANCOVA (F_{NA} (1, 34) = 2.93, $p = .09$, $\eta^2_{partial} = .08$; *cf. Figure 8 in the supplementary material section*).

Valence and Arousal. Analysis of covariance revealed no main effect of aesthetic appeal on valence and arousal ($F_{valence}$ (1, 35) < 1, $\eta^2_{partial} = .01$; $F_{arousal}$ (1, 35) < 1, $\eta^2_{partial} = .02$). These results indicate that the WIBIG hypothesis could not be confirmed with regard to the influence of gustatory aesthetics on measures of affect in the context of UX evaluations (*cf. Figure 8 in the supplementary material section*).

User efficiency. The analysis of the overall task completion time showed no difference between means as a function of gustatory aesthetics (t (35) = 0.28; $p = .39$, $d = 0.09$). In a similar vein, the number of user interactions did not vary according to the manipulation of gustatory aesthetics (t (36) = 0.50; $p = .31$, $d = 0.16$). Based on these results, the WIBIG hypothesis must be rejected regarding the influence of gustatory aesthetics on user efficiency (*cf. Figure 9 in the supplementary material section*).

CORRELATIONAL ANALYSIS OF UX MEASURES

In order to better understand the interplay between the different components of the UX construct, analysis of correlations is generally considered to be very useful (cf. Hornbæk & Law, 2007; Sonderegger et al., 2019). Correlations between the components of the UX construct assessed in the three studies are summarised in Table 2. As expected due to theoretical deliberations, measures of affect showed strong correlations among each other while the link to arousal was small. In a similar vein, the negative association of perceived usability with task completion time (though not always significant; r ranging between –.26 and –.35) indicates the link between objective and subjective usability measures. The rather large effects of perceived usability with measures of affect (for PA r ranging between .29 and .51, for NA r ranging between –.05 and –.64, for valence r ranging between .08 and .59) hint at the relevance of the affect mediation model (Norman, 2002). Perceived gustatory attractiveness showed rather small correlations with other UX measures, but in study III, some correlations with measures of affect reached significance levels (r = .41 for valence and r = –.36 for NA). Overall, the correlational analysis indicated that the different components of the UX construct correlate weakly with each other, which suggests that in the context of UX evaluations, the constructs should be assessed and reported separately rather than an overall UX score being computed (Sonderegger et al., 2019).

DISCUSSION

Despite its small sample size (post-hoc analyses revealed for perceived usability a statistical power

MEASURES	PERCEIVED GUSTATORY ATTRACTIVENESS	POSITIVE AFFECT	NEGATIVE AFFECT	VALENCE (SAM)	AROUSAL (SAM)	EFFICIENCY (TASK COMPLETION TIME)
Perceived usability						
Study I	**.33 (.04)**	**.34 (.03)**	–.05 (.75)	.08 (.61)	–.04 (.81)	–.26 (.10)
Study II	.13 (.31)	**.51 (.00)**	**–.57 (.00)**	**.59 (.00)**	–.14 (.28)	**–.35 (.01)**
Study III	**.31 (.05)**	.29 (.08)	**–.64 (.00)**	**.44 (.01)**	.11 (.53)	–.29 (.08)
Perceived gustatory attractiveness						
Study I		.17 (.30)	–.13 (.44)	.26 (.11)	–.11 (.49)	–.05 (.74)
Study II		.15 (.25)	.08 (.54)	.14 (.28)	–.03 (.83)	.07 (.59)
Study III		.08 (.64)	**–.36 (.03)**	**.41 (.01)**	.16 (.34)	.02 (.90)
Positive Affect						
Study I			.13 (.44)	.25 (.13)	.01 (.94)	–.10 (.53)
Study II			**–.33 (.01)**	**.63 (.00)**	–.08 (.57)	**–.33 (.01)**
Study III			–.14 (.40)	.25 (.14)	.32 (.05)	–.29 (.08)
Negative Affect						
Study I				–30. (.05)	**.37 (.02)**	.15 (.36)
Study II				**–.43 (.01)**	**.31 (.01)**	.17 (.20)
Study III				**–54. (.00)**	.11 (.52)	.28 (.09)
Valence (SAM)						
Study I					–.12 (.46)	.31 (.05)
Study II					.04 (.76)	**–.26 (.05)**
Study III					–.08 (.65)	–.11 (.52)
Arousal (SAM)						
Study I						.04 (.80)
Study II						.11 (.39)
Study III						–.14 (.41)

Table 2 Correlations (two-tailed, with exact p-value in parentheses) between dependent variables in the three studies (N_{StudyI} = 40, $N_{StudyII}$ = 60, $N_{StudyIII}$ = 38).

Bold numbers denote significant correlation coefficients in order to make them better distinguishable from non-significant correlations. The shading represents the magnitude of the average correlation coefficients based on Cohen's effect size classification (white = no effect, light grey = small effect, grey = medium effect, and dark grey = large effect).

of .08 (α = .05), indicating a very high risk of Type II errors), the results of the third experiment are in line with the findings of the two preceding ones, indicating that the WIBIG effect for aesthetics of taste on various UX measures is considerably smaller than the effect for visual aesthetics reported in previous research. However, the small effect size may also be partially due to the specific sample of self-selected coffee experts, who might possess rigid, well-defined criteria for coffee evaluation, making them less susceptible to cognitive biases such as the halo effect in comparison to average consumers.

An important additional result of this third study however is of methodological nature: It appears to be easier to manipulate coffee quality if it involves making a comparative judgement rather than an absolute judgement using a 'contrast coffee'. This was indicated by a very large effect for the manipulation check, with the effect size being larger than in the two previous studies. This suggests that using comparative judgements rather than absolute ones in the methodological approach may help make manipulations of aesthetics in experimental research more effective.

GENERAL DISCUSSION

The primary research question examined whether a halo effect of aesthetics of taste on UX could be found, with effect sizes comparable to those of visual aesthetics reported in previous research. However, none of the three experiments showed any evidence for similar effects as shown for visual aesthetics. The findings suggest that aesthetics of taste produces a similar picture as aesthetics of sound and touch did in previous work (Sonderegger & Sauer, 2015). This indicates that the effect of non-visual aesthetics on UX measures is smaller compared to the impact of visual aesthetics.

There are several possible reasons that may account for this apparent fundamental difference between visual aesthetics and non-visual aesthetics (the reader is also referred to the discussion in Sonderegger & Sauer, 2015): (a) Experience of users may be less influenced by the gustatory outcome quality of the system than by its visual appearance. (b) It may be related to the observation that the visual modality is able to process product features very quickly (e.g., Lindgaard et al., 2006; Schifferstein & Desmet, 2007), whereas such information processing might take more time for non-visual modalities. (c) Cognitive biases such as the halo effect were thought to explain the 'what is beautiful is good'-principle (e.g., Tractinsky et al., 2000), but they may be less likely to work in a non-visual context because non-visual features are in general less salient than visual features. (d) The issue of sensory dominance (Fenko et al., 2010; Schifferstein, 2006) may have played a role here since it suggests very different time-based patterns for each sensory modality. Fenko et al. (2010) have argued that the importance of the visual modality decreases over time while non-visual modalities show the opposite pattern. With most UX research relying on a one-off testing session, it suggests that for methodological reasons the influence of the visual modality may have been overestimated.

It is also important to note that there was no corroborated interaction between aesthetics of taste and system usability. This is important since previous research indicated that visual aesthetics showed mainly effects on user behaviour in situations where poor usability was modelled (Reppa & McDougall, 2015; Moshagen et al., 2009). In addition to the usability manipulation, we also used changes in the methodological approach (e.g., different levels of coffee expertise of participants, different types of coffee, tasks, comparative judgement of coffee rather than absolute judgement) to determine whether the effects were consistent across the different conditions. Since they were found to be, it further strengthened our argument for the existence of different result patterns for visual and gustatory aesthetics.

Although we were referring to the attribution theory (Fiske & Taylor, 1991; Weiner, 1985) when formulating our hypotheses, the presented result pattern could be linked to the Dual Process Theory (e.g., Kahneman & Frederick, 2002; Wason & Evans, 1974) suggesting that two competing cognitive systems, System 1 (fast, automatic, and intuitively processing) and System 2 (slow, deliberate, and analytical processing) influence decision making. This framework provides potential hints for the missing link between gustatory aesthetics and UX measures by suggesting that immediate sensory impressions of taste engage System 1, while UX evaluations rely on System 2. This effect pattern could be explained within the framework of processing fluency theory (Reber et al., 2004; see also Belke et al., 2015; Douneva et al., 2016; McDougall et al., 2016; Orth & Wirtz, 2014), suggesting that when a visual stimulus is fast and easy to process, it is perceived as being aesthetically pleasing. This fluent processing is closely tied to a user's experience: what is easy to process is also easy to use, which could explain why visual aesthetics are strongly linked with usability judgments. In contrast, processing fluency play a smaller role in gustatory experience and therefore be less directly linked to measures of UX.

The inconsistent effect of aesthetics of taste on affect cannot be interpreted conclusively based on the obtained data. While the nil effects of experiment I might be due to the small effect of the gustatory aesthetics manipulation, the small effect of experiments II and III may be considered a surprising result even if it is largely in line with other work examining the relation between non-visual aesthetics and affective states (Sonderegger & Sauer, 2015). Non-visual modalities are generally associated with more intense emotions than vision (Sonneveld & Schifferstein, 2008), with empirical work on visual aesthetics showing effects on affect (e.g., Thüring & Mahlke, 2007). Furthermore, it has been argued that the aesthetic experience is a process characterised by high levels of affect (Leder et al., 2004). This has been shown in experiment II, where aesthetics of taste showed a small positive influence on valence. This small effect might be linked to the manipulation of aesthetics of taste in that study (i.e. presentation of the most or least preferred coffee), whereby the fact of receiving the chosen option or not may have had a decisive influence on the affective state. Furthermore, the relatively low internal consistency observed for the Negative Affect (NA) subscale of the I-PANAS-SF needs to be taken into consideration. Low reliability increases measurement error and can attenuate observed relationships between constructs, thereby reducing statistical power to detect existing effects. Consequently, the absence of significant associations between aesthetics of taste and negative affect should be interpreted with caution, as any true relationships may have been underestimated. At the same time, the relatively consistent pattern of results across experiments suggests that the findings are not solely attributable to measurement unreliability. Further research is therefore needed to better understand the link between aesthetics of taste and affect.

Regarding the limitations of this study, it is acknowledged that only one type of product was examined. It will be interesting to see whether future research examining other systems producing gustatory experiences than coffee machines will obtain similar results as the present three experiments. As a further limitation, aesthetics of taste was not related to the device itself but to the central outcome of the device. This represents a methodological problem which is linked to the fact that the main function of the modality of taste is related to users consuming food (Schifferstein & Spence, 2008). Nevertheless, the

assumption that this outcome of an interaction with the system (e.g., taste of coffee) influences the experience of the interaction (i.e. UX linked with the use of the coffee machine) does not seem unreasonable based on theoretical assumptions following attribution theory (see introduction). We furthermore acknowledge that the usability manipulation in experiments II and III (i.e. power failure and impossible task) may be considered as only partially matching the core of the usability concept since it is strongly related to the concept of system reliability. However, such a manipulation offered the advantage that it was difficult to compensate by participants (e.g., by finding a way to work round it) and resulted in comparable effect sizes to the taste manipulation. Furthermore, reliability is generally considered a system-inherent quality attribute, as it refers to the extent to which a system performs consistently and without failure, while usability is typically regarded as a user-centred attribute, as it reflects users' perceptions of how effective, efficient, and satisfying a system is to use. As users often infer a system's usability from its performance characteristics, manipulating system reliability can be viewed as a valid means of influencing users' subjective experience of usability. For example, frequent delays, errors, or system failures may lead users to perceive a system as more difficult to use, even if the interface itself remains unchanged. As a further limitation, the studies used one-off usability tests. There is however some empirical evidence that the effect of visual aesthetics on usability measures diminishes over time (e.g., Sonderegger et al., 2012) and that other sensory modalities gain in importance with prolonged usage (Fenko et al., 2010). Although one-off usability evaluations are commonly applied in practice, such results indicate the need for longitudinal research observing the respective effects of visual and gustatory aesthetics over the course of time. Based on the results of previous studies as well as the findings of this piece of research, an interaction between the two factors could be assumed, with visual aesthetics initially showing a strong effect on measures of UX which wanes over time while aesthetics of taste becomes increasingly important. A further limitation may be that smell and taste of the coffee have not been conceptualised and assessed separately in this study. Although it is generally agreed upon that they are considerably linked (Castillo, 2014; Schifferstein et al., 2022; Small & Prescott, 2005; Spence, 2015), a specific distinction between olfactory and gustatory experience might be interesting for future studies. In a similar vein, future studies might benefit from employing multi-item scales that capture the diverse aspects of gustatory evaluation and provide a more comprehensive assessment of the construct. Although single-item measures offer practical advantages, such as reduced participant burden and shorter survey completion times, they may not adequately capture the multifaceted nature of gustatory experiences. Finally, it must be acknowledged that the samples in all three studies were too small to be able to make a confident claim regarding the absence of a WIBIG effect. However, the results of the three studies presented above might be interpreted as a first indicator that the role gustatory aesthetics play in UX might be of lower importance than the one of visual aesthetics, which is a finding that needs to be replicated in future studies and meta-analyses. In this regard, these results represent an important piece of a jigsaw for a better understanding of the influence of non-visual aesthetics on user experience.

Being the first work that examined the role of aesthetics of taste in UX evaluation, it provided first evidence for a different pattern of outcomes compared to visual aesthetics. At the same time, it provided more evidence for considerable similarity between different types of non-visual aesthetics. As a complementary remark, we would like to point out that these considerations apply to users without visual impairments. In the context of accessibility research (Vanderheiden & Jordan, 2012), it has to be considered that for users with visual impairments, the order of importance of sensory modalities is a completely different one. It would therefore be interesting to examine whether for users with visual impairments an effect of non-visual aesthetics on user experience could be demonstrated.

The present studies were also characterised by a continuous effort to improve the manipulation of aesthetics of taste to achieve the largest possible effects. The obtained increase in effect size in the manipulation checks is an interesting finding, which may be worth examining. It suggests considerable interpersonal differences regarding the gustatory attractiveness of products (e.g., participants usually drinking cappuccino or tea did not differentiate between good and bad espresso in our first study). This makes the group-based approach to manipulating aesthetics of taste less promising than the person-based approach relying on individual preferences (e.g., adding milk and/or sugar as adopted in study II). However, by using the contrast effect in the form of offering a 'contrast coffee', the group-based approach gained in effectiveness, which was confirmed by the findings of the pilot studies in experiments I and III. While none of the between-subjects comparisons (i.e. no contrast present) revealed a difference between the good and the bad coffee, the repeated-measures designs (i.e. with a contrast coffee) showed large effects of our manipulation of coffee taste. The presentation of a contrast stimulus (i.e. the possibility for test participants to compare two versions of the coffee) hence proved to be an important factor for a successful manipulation of aesthetics of taste. For UX practitioners conducting A/B

testing with an interest in effects of aesthetics of taste, the findings of the studies presented in this article hint at advantages of manipulating gustatory aesthetics using contrast stimuli because between-subjects manipulations might not lead to effects of sufficient magnitude.

The challenges in the manipulation of aesthetics of taste in the present series of studies points to a fundamental problem in the domain of experimental research on aesthetics (Tractinsky, 2013). Is a group-based approach to manipulating aesthetics useful or is the aesthetic experience a highly personal matter ('tastes are different' or 'beauty is in the eye of the beholder') and therefore needs to be addressed on a personal level? Only a small number of studies have been published that succeeded in the experimental manipulation of (visual) artefact attractiveness using a group-based approach (e.g., Sonderegger et al., 2012; Sonderegger & Sauer, 2010; Tuch et al., 2012). Our own research has shown that a group-based manipulation of visual aesthetics often failed as an experimental manipulation. Such work is usually not published. It might be due to a publication bias that only a small number of studies are available, in which a group-based manipulation of aesthetics was applied. Interestingly, most published studies that manipulated visual aesthetics used a repeated-measures design, in which participants interacted with both versions of the artefact, that is, the aesthetically pleasing and the unpleasing one (e.g., Ben-Bassat et al., 2006; Thüring & Mahlke, 2007). This allowed participants to compare the two types of design, which is similar to the contrast effect used in our work. The results presented in this series of studies indicate that the manipulation of aesthetics can also be successful in a between-subjects design. Presenting a contrast stimulus might increase the strength of the experimental manipulation of aesthetics. The approach that was used to manipulate aesthetics of taste (preference approach, contrast approach) could hence also be applied to the manipulation of aesthetics in other sensory modalities to overcome problems of achieving a successful manipulation.

CONCLUSION

In addition to the intriguing methodological findings regarding experimental manipulation of gustatory aesthetics, the results presented in this paper indicated that in stark contrast to empirical results on the influence of visual aesthetics, the effect of taste of the outcome product is smaller in UX evaluations. If future studies with greater statistical power and longitudinal designs corroborate these results, our findings could be considered rather surprising, considering the importance of taste in food producing products.

TRANSPARENCY STATEMENT

We reported how we determined the sample size and the stopping criterion. We reported all experimental conditions and variables. We reported all data exclusion criteria and whether these were determined before or during the data analysis. We reported all outlier criteria and whether these were determined before or during data analysis.

NOTES

1 In the present study, aesthetics, or aesthetic appeal, is defined as an objective characteristic of the product, whereas the individual's response to and subjective evaluation of this characteristic is referred to as attractiveness (Sonderegger & Sauer, 2010, see also Moshagen & Thielsch, 2010).

2 All data generated and analyzed in the three studies are available on the Open Science Framework, https://doi.org/10.17605/OSF.IO/2KHFR.

The research plans of the three studies have not been pre-registered prior to the research being conducted.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Supplemental Material.** Result section for experiment II. DOI: https://doi.org/10.5334/spo.94.s1

ACKNOWLEDGEMENTS

We are grateful to Kim Noémie Buchmüller, Yusuf Arslanparcasi and Larissa Berchtold for their help with data collection. We would also like to thank Jura Ltd. (Switzerland) for providing us with two coffee machines for this research project.

FUNDING INFORMATION

The work was financially supported by a research grant (No 100014_140359) from the Swiss National Science Foundation (SNSF).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualisation, data curation, formal analysis, writing original draft, methodology, supervision, project administration. Author 2: Conceptualisation, formal analysis, writing original draft, investigation. Author 3: Conceptualisation, writing original draft, methodology, supervision, funding acquisition.

AUTHOR AFFILIATIONS

Andreas Sonderegger orcid.org/0000-0003-0054-0544
FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
Juergen Baumgartner orcid.org/0000-0003-1341-7502
University of Fribourg, Switzerland
Juergen Sauer orcid.org/0000-0003-2105-1694
University of Fribourg, Switzerland

REFERENCES

Allen, M. S., Iliescu, D., & **Greiff, S.** (2022). Single Item Measures in Psychological Science: A Call to Action. *European Journal of Psychological Assessment*, *38*(1). https://doi.org/10.1027/1015-5759/a000699

Anderson, S. F. (2019). Best (but oft forgotten) practices: Sample size planning for powerful studies. *The American Journal of Clinical Nutrition*, *110*(2), 280–295. https://doi.org/10.1093/ajcn/nqz058

Assila, A., Oliveira, K. M. de, & **Ezzedine, H.** (2016). Standardized Usability Questionnaires: Features and Quality Focus. *Electronic Journal of Computer Science and Information Technology*, *6*(1). https://ejcsit.uniten.edu.my/index.php/ejcsit/article/view/96

Bangor, A., Kortum, P. T., & **Miller, J. T.** (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barber, R. E., & **Lucas, H. C.** (1983). System response time operator productivity, and job satisfaction. *Communications of the ACM*, *26*(11), 972–986. https://doi.org/10.1145/182.358464

Belke, B., Leder, H., & **Carbon, C. C.** (2015). When Challenging Art Gets Liked: Evidences for a Dual Preference Formation Process for Fluent and Non-Fluent Portraits. *PLoS ONE*, *10*(8), e0131796. https://doi.org/10.1371/journal.pone.0131796

Ben-Bassat, T., Meyer, J., & **Tractinsky, N.** (2006). Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transactions on Computer-Human Interaction*, *13*(2), 210–234. https://doi.org/10.1145/1165734.1165737

Bradley, M. M., & **Lang, P. J.** (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brooke, J. (1996). SUS-A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability Evaluation In Industry* (pp. 189–194). Taylor&Francis.

Brun, L., Pansu, P., & **Dompnier, B.** (2021). The role of causal attributions in determining behavioral consequences: A meta-analysis from an intrapersonal attributional perspective in achievement contexts. *Psychological Bulletin*, *147*(7), 701–718. https://doi.org/10.1037/bul0000331

Cameron, M. I., Morisco, D., Hofstetter, D., Uman, E., Wilkinson, J., Kennedy, Z. C., Fontenot, S. A., Lee, W. T., Hendon, C. H., & **Foster, J. M.** (2020). Systematically Improving Espresso: Insights from Mathematical Modeling and Experiment. *Matter*, *2*(3), 631–648. https://doi.org/10.1016/j.matt.2019.12.019

Cardello, A. V., & **Wise, P. M.** (2008). Taste, Smell and Chemesthesis in Product Experience. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product Experience* (pp. 91–131). Elsevier. https://doi.org/10.1016/B978-008045089-6.50007-1

Castillo, M. (Ed.) (2014). The Complicated Equation of Smell, Flavor, and Taste. *American Journal of Neuroradiology*, *35*(7), 1243–1245. https://doi.org/10.3174/ajnr.A3739

Chawda, B., Craft, B., Cairns, P., Heesch, D., & **Rüger, S.** (2005). Do 'attractive things work better'? An exploration of search tool visualisations. *19th British HCI Group Annual Conference: The Bigger Picture*, 46–51. https://oro.open.ac.uk/29873/

Chen, C., & **He, G.** (2016). The Contrast Effect in Temporal and Probabilistic Discounting. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00304

De Angeli, A., Sutcliffe, A., & **Hartmann, J.** (2006). Interaction, usability and aesthetics: What influences users' preferences? *Proceedings of the 6th Conference on Designing Interactive Systems*, 271–280. https://dl.acm.org/citation.cfm?id=1142446

Desmet, P. M. A., Hekkert, P., & **Jacobs, J. J.** (2000). When a Car Makes You Smile: Development and Application of an Instrument to Measure Product Emotions. *Advances in Consumer Research*, *27*(1), 111–117. (6687243).

Dewey, J. (2005). *Art as Experience*. Penguin.

Dion, K., Berscheid, E., & **Walster, E.** (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. https://doi.org/10.1037/h0033731

Douneva, M., Jaron, R., & **Thielsch, M. T.** (2016). Effects of Different Website Designs on First Impressions, Aesthetic Judgements and Memory Performance after Short Presentation. *Interacting with Computers*, *28*(4), 552–567. https://doi.org/10.1093/iwc/iwv033

Fakunmoju, S. B. (2020). Validity of Single-item Versus Multiple-item Job Satisfaction Measures in Predicting Life: Satisfaction and Turnover Intention. *Asia-Pacific Journal of Management Research and Innovation*, *16*(3), 210–228. https://doi.org/10.1177/2319510X21997724

Faul, F., Erdfelder, E., Lang, A.-G., & **Buchner, A.** (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fenko, A., Schifferstein, H. N. J., & **Hekkert, P.** (2010). Shifts in sensory dominance between various stages of user–product interactions. *Applied Ergonomics*, *41*(1), 34–40. https://doi.org/10.1016/j.apergo.2009.03.007

Fisicaro, S. A., & **Lance, C. E.** (1990). Implications of Three Causal Models for the Measurement of Halo Error. *Applied Psychological Measurement*, *14*(4), 419–429. https://doi.org/10.1177/014662169001400407

Fiske, S. T., & **Taylor, S. E.** (1991). *Social cognition*. Mcgraw-Hill. https://psycnet.apa.org/record/1991-97723-000

Folkes, V. S. (1988). Recent Attribution Research in Consumer Behavior: A Review and New Directions. *Journal of Consumer Research*, *14*(4), 548–565. https://doi.org/10.1086/209135

Galletta, D. F., Henry, R., McCoy, S., & **Polak, P.** (2004). Web Site Delays: How Tolerant are Users? *Journal of the Association for Information Systems*, *5*(1), Article 1. https://doi.org/10.17705/1jais.00044

Gayler, T., & **Sas, C.** (2017). An exploration of taste-emotion mappings from the perspective of food design practitioners. *Proceedings of the 2nd ACM SIGCHI International Workshop on Multisensory Approaches to Human-Food Interaction, MHFI 2017*, 23–28. https://doi.org/10.1145/3141788.3141793

Gayler, T., Sas, C., & **Kalnikaite, V.** (2019). Taste Your Emotions: An Exploration of the Relationship between Taste and Emotional Experience for HCI. *Proceedings of the 2019 on Designing Interactive Systems Conference, DIS '19*, 1279–1291. https://doi.org/10.1145/3322276.3322336

Gibson, J. J. (1966). *The senses considered as perceptual systems*. GEORGE ALLEN & UNWIN LTD.

Giraldo-O'Meara, M., Marin-Garcia, J. A., & **Martinez-Gomez, M.** (2014). Validation of the JDS satisfaction scales applied to educational university environments. *Journal of Industrial Engineering and Management*, *7*(1), 72–99. https://doi.org/10.3926/jiem.906

Harmon-Jones, C., Bastian, B., & **Harmon-Jones, E.** (2016). The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-Reported Emotions. *PLOS ONE*, *11*(8), e0159915. https://doi.org/10.1371/journal.pone.0159915

Hartmann, J., Sutcliffe, A., & **De Angeli, A.** (2007). Investigating attractiveness in web user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, 387–396. https://doi.org/10.1145/1240624.1240687

Hassenzahl, M., & **Monk, A.** (2010). The inference of perceived usability from beauty. *Human–Computer Interaction*, *25*(3), 235–260. https://doi.org/10.1080/07370024.2010.500139

Hekkert, P., & **Leder, H.** (2008). Product aesthetics. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product Experience* (pp. 259–285). Elsevier. https://doi.org/10.1016/B978-008045089-6.50013-7

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Hornbæk, K., & **Law, E. L.-C.** (2007). Meta-analysis of Correlations Among Usability Measures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, 617–626. https://doi.org/10.1145/1240624.1240722

ISO 9241-210. (2019). *Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*. International Organization for Standardization. https://www.iso.org/standard/77520.html

JASP Team. (2022). *JASP (version 0.16.3)[Computer software]* [Computer software]. https://jasp-stats.org/

Jordan, P. W. (2020). *An Introduction to Usability*. CRC Press. https://doi.org/10.1201/9781003062769

Kahneman, D., & **Frederick, S.** (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & **Tausch, A.** (1996). Untersuchungen mit einer deutschen version der "positive and negative affect schedule" (PANAS). *Diagnostica-Gottingen-*, *42*, 139–156. https://doi.org/10.1037/t49650-000

Kurosu, M., & **Kashimura, K.** (1995). Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. *Conference Companion on Human Factors in Computing Systems*, 292–293. https://doi.org/10.1145/223355.223680

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & **Isager, P. M.** (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Law, E. L.-C., & **van Schaik, P.** (2010). Modelling user experience – An agenda for research and practice. *Interacting with Computers*, *22*(5), 313–322. https://doi.org/10.1016/j.intcom.2010.04.006

Leder, H., Belke, B., Oeberst, A., & **Augustin, D.** (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. https://doi.org/10.1348/0007126042369811

Lenhard, W., & **Lenhard, A.** (2016). *Berechnung von Effektstärken*. https://www.psychometrica.de/effektstaerke.html

Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Lindgaard, G., Fernandes, G., Dudek, C., & **Brown, J.** (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, *25*(2), 115–126. https://doi.org/10.1080/01449290500330448

McDougall, S. J., Reppa, I., Kulik, J., & **Taylor, A.** (2016). What makes icons appealing? The role of processing fluency in predicting icon appeal in different task contexts. *Applied Ergonomics*, *55*, 156–172. https://doi.org/10.1016/j.apergo.2016.02.006

Mehrabian, A., & **Russell, J. A.** (1974). *An approach to environmental psychology*. the MIT Press.

Miller, C. (2011). Aesthetics and e-assessment: The interplay of emotional design and learner performance. *Distance Education*, *32*(3), 307–337. https://doi.org/10.1080/01587919.2011.610291

Moshagen, M., Musch, J., & **Göritz, A. S.** (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, *52*(10), 1311–1320. https://doi.org/10.1080/00140130903061717

Moshagen, M., & **Thielsch, M. T.** (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, *68*(10), 689–709. https://doi.org/10.1016/j.ijhcs.2010.05.006

Murer, M., Aslan, I., & **Tscheligi, M.** (2013). LOLL*io*: Exploring taste as playful modality. *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, TEI '13*, 299–302. https://doi.org/10.1145/2460625.2460675

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, *75*(1), 77–86. https://doi.org/10.1348/096317902167658

Norman, D. (2002). Emotion & Design: Attractive Things Work Better. *Interactions*, *9*(4), 36–42. https://doi.org/10.1145/543434.543435

Obrist, M., Comber, R., Subramanian, S., Piqueras-Fiszman, B., Velasco, C., & **Spence, C.** (2014). Temporal, affective, and embodied characteristics of taste experiences: A framework for design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, 2853–2862. https://doi.org/10.1145/2556288.2557007

Orth, U. R., & **Wirtz, J.** (2014). Consumer Processing of Interior Service Environments The Interplay Among Visual Complexity, Processing Fluency, and Attractiveness. *Journal of Service Research*, 1094670514529606. https://doi.org/10.1177/1094670514529606

Perullo, N. (2016). *Taste as Experience: The Philosophy and Aesthetics of Food*. Columbia University Press. https://doi.org/10.7312/peru17348

Reber, R., Schwarz, N., & **Winkielman, P.** (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Reppa, I., & **McDougall, S. J.** (2015). When the going gets tough the beautiful get going: Aesthetic appeal facilitates task performance. *Psychonomic Bulletin & Review*, *22*(5), 1243–1254. https://doi.org/10.3758/s13423-014-0794-z

Reppa, I., McDougall, S., Sonderegger, A., & **Schmidt, W. C.** (2021). Mood moderates the effect of aesthetic appeal on performance. *Cognition and Emotion*, *35*(1), 15–29. https://doi.org/10.1080/02699931.2020.1800446

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

Russell, J. A., & **Barrett, L. F.** (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Sander, D. (2013). Models of emotion. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge Handbook of Human Affective Neuroscience* (pp. 5–53). Cambridge University Press.

Sauer, J., & **Sonderegger, A.** (2009). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics*, *40*(4), 670–677. https://doi.org/10.1016/j.apergo.2008.06.006

Sauer, J., & **Sonderegger, A.** (2022). Visual aesthetics and user experience: A multiple-session experiment. *International Journal of Human-Computer Studies*, *165*, 102837. https://doi.org/10.1016/j.ijhcs.2022.102837

Sauer, J., Sonderegger, A., & **Schmutz, S.** (2020). Usability, user experience and accessibility: Towards an integrative model. *Ergonomics*, *63*(10), 1207–1220. https://doi.org/10.1080/00140139.2020.1774080

Sauro, J., & **Lewis, J. R.** (2012). *Quantifying the User Experience: Practical Statistics for User Research* (1st edn). Morgan Kaufmann Publishers Inc. https://doi.org/10.1016/B978-0-12-384968-7.00002-3

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R., Shuman, V., Fontaine, J. R., & **Soriano, C.** (2013). The GRID meets the Wheel: Assessing emotional feeling via self-report. In K. R. Scherer & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281–298). Oxford University Press.

Schifferstein, H. (2006). The perceived importance of sensory modalities in product usage: A study of self-reports. *Acta Psychologica*, *121*(1), 41–64. https://doi.org/10.1016/j.actpsy.2005.06.004

Schifferstein, H., & **Cleiren, M.** (2005). Capturing product experiences: A split-modality approach. *Acta Psychologica*, *118*(3), 293–318. https://doi.org/10.1016/j.actpsy.2004.10.009

Schifferstein, H., & **Desmet, P. M. A.** (2007). The effects of sensory impairments on product experience and personal well-being. *Ergonomics*, *50*(12), 2026–2048. https://doi.org/10.1080/00140130701524056

Schifferstein, H., Otten, J. J., Thoolen, F., & **Hekkert, P.** (2010). The experimental assessment of sensory dominance in a product development context. *Journal of Design Research*, *8*(2), 119–144. https://doi.org/10.1504/JDR.2010.032074

Schifferstein, H., & **Spence, C.** (2008). Multisensory Product Experience. In H. Schifferstein & P. Hekkert (Eds.), *Product experience* (pp. 133–161). Elsevier. https://doi.org/10.1016/B978-008045089-6.50008-3

Schifferstein, H. N. J., Kudrowitz, B. M., & **Breuer, C.** (2022). Food Perception and Aesthetics—Linking Sensory Science to Culinary Practice. *Journal of Culinary Science & Technology*, *20*(4), 293–335. https://doi.org/10.1080/15428052.2020.1824833

Seo, K.-K., Lee, S., & **Chung, B. D.** (2016). Effects of Perceived Usability and Aesthetics on Emotional Responses in Different Contexts of Use. *International Journal of Human–Computer Interaction*, *32*(6), 445–459. https://doi.org/10.1080/10447318.2016.1160537

Small, D. M., & **Prescott, J.** (2005). Odor/taste integration and the perception of flavor. *Experimental Brain Research*, *166*(3), 345–357. https://doi.org/10.1007/s00221-005-2376-9

Sonderegger, A., & **Sauer, J.** (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, *41*(3), 403–410. https://doi.org/10.1016/j.apergo.2009.09.002

Sonderegger, A., & **Sauer, J.** (2015). The role of non-visual aesthetics in consumer product evaluation. *International Journal of Human-Computer Studies*, *84*, 19–32. https://doi.org/10.1016/j.ijhcs.2015.05.011

Sonderegger, A., Sauer, J., & **Eichenberger, J.** (2014). Expressive and classical aesthetics: Two distinct concepts with highly similar effect patterns in user–artefact interaction. *Behaviour & Information Technology*, *33*(11), 1180–1191. https://doi.org/10.1080/0144929X.2013.853835

Sonderegger, A., Uebelbacher, A., Pugliese, M., & **Sauer, J.** (2014). The Influence of Aesthetics in Usability Testing: The Case of Dual-domain Products. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, 21–30. https://doi.org/10.1145/2556288.2557419

Sonderegger, A., Uebelbacher, A., & **Sauer, J.** (2019). The UX Construct – Does the Usage Context Influence the Outcome of User Experience Evaluations? In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer Interaction – INTERACT 2019* (pp. 140–157). Springer International Publishing. https://doi.org/10.1007/978-3-030-29390-1_8

Sonderegger, A., Zbinden, G., Uebelbacher, A., & **Sauer, J.** (2012). The influence of product aesthetics and usability over the course of time: A longitudinal field experiment. *Ergonomics*, *55*(7), 713–730. https://doi.org/10.1080/00140139.2012.672658

Sonneveld, M. H., & **Schifferstein, H. N. J.** (2008). The tactual experience of objects. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product Experience* (pp. 41–67). Elsevier. https://doi.org/10.1016/B978-008045089-6.50005-8

Spence, C. (2015). Just how much of what we taste derives from the sense of smell? *Flavour*, *4*(1), 30. https://doi.org/10.1186/s13411-015-0040-2

Szameitat, A. J., Rummel, J., Szameitat, D. P., & **Sterr, A.** (2009). Behavioral and emotional consequences of brief delays in human–computer interaction. *International Journal of Human-Computer Studies*, *67*(7), 561–570. https://doi.org/10.1016/j.ijhcs.2009.02.004

Thielsch, M. T., Blotenberg, I., & **Jaron, R.** (2014). User Evaluation of Websites: From First Impression to Recommendation. *Interacting with Computers*, *26*(1), 89–102. https://doi.org/10.1093/iwc/iwt033

Thielsch, M. T., Engel, R., & **Hirschfeld, G.** (2015). Expected usability is not a valid indicator of experienced usability. *PeerJ Computer Science*, *1*, e19. https://doi.org/10.7717/peerj-cs.19

Thielsch, M. T., Scharfen, J., Masoudi, E., & **Reuter, M.** (2019). Visual Aesthetics and Performance: A First Meta-Analysis. *Proceedings of Mensch Und Computer 2019, MuC'19*, 199–210. https://doi.org/10.1145/3340764.3340794

Thompson, E. R. (2007). Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. https://doi.org/10.1177/0022022106297301

Thüring, M., & **Mahlke, S.** (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, *42*(4), 253–264. https://doi.org/10.1080/00207590701396674

Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 115–122. https://dl.acm.org/citation.cfm?id=258626

Tractinsky, N. (2013). Visual Aesthetics. In *The Encyclopedia of Human-Computer Interaction, 2nd Ed, 19*. Interaction Design Foundation.

Tractinsky, N., Katz, A. S., & **Ikar, D.** (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145. https://doi.org/10.1016/S0953-5438(00)00031-X

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & **Wilhelm, F. H.** (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, *67*(9), 703–715. https://doi.org/10.1016/j.ijhcs.2009.04.002

Tuch, A. N., Roth, S. P., Hornbæk, K., Opwis, K., & **Bargas-Avila, J. A.** (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior*, *28*(5), 1596–1607. https://doi.org/10.1016/j.chb.2012.03.024

Vanderheiden, G. C., & **Jordan, J. B.** (2012). Design for People with Functional Limitations. In *Handbook of Human Factors and Ergonomics* (pp. 1407–1441). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118131350.ch51

Wason, P. C., & **Evans, J. ST. B. T.** (1974). Dual processes in reasoning? *Cognition*, *3*(2), 141–154. https://doi.org/10.1016/0010-0277(74)90017-1

Watson, D., Clark, L. A., & **Tellegen, A.** (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weiner, B. (1985). Attribution Theory. In B. Weiner (Ed.), *Human Motivation* (pp. 275–326). Springer. https://doi.org/10.1007/978-1-4612-5092-0_7

Weiner, B. (2000). Attributional Thoughts about Consumer Behavior. *Journal of Consumer Research*, *27*(3), 382–387. https://doi.org/10.1086/317592

Wright, P., McCarthy, J., & **Meekison, L.** (2003). Making Sense of Experience. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology* (pp. 43–53). Springer Netherlands. https://doi.org/10.1007/1-4020-2967-5_5

Yusof, U. K., Khaw, L. K., Ch'ng, H. Y., & **Neow, B. J.** (2010). Balancing between usability and aesthetics of Web design. *2010 International Symposium on Information Technology*, *1*, 1–6. https://doi.org/10.1109/ITSIM.2010.5561310

TO CITE THIS ARTICLE:
Sonderegger, A., Baumgartner, J., & Sauer, J. (2026). Aesthetics of Taste in Evaluations of User Experience. *Swiss Psychology Open*, 6(1): 8, pp. 1–23. DOI: https://doi.org/10.5334/spo.94

Submitted: 11 July 2025 **Accepted:** 02 November 2026 **Published:** 25 November 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Swiss Psychology Open is a peer-reviewed open access journal published by Ubiquity Press.