Temporal Dominance of Sensations paired with dynamic wanting in an ad libitum setting: A new method of sensory evaluation with consumers for a better understanding of beer drinkability

Takahiro Wakihira, Seiko Miyashita, Minoru Kobayashi, Kazuhiko Uemura, Pascal Schlich

[Title]

Temporal Dominance of Sensations paired with dynamic wanting in an *ad libitum* setting: a new method of sensory evaluation with consumers for a better understanding of beer drinkability

[Authors]

Takahiro Wakihira[a], Seiko Miyashita[a], Minoru Kobayashi[a], Kazuhiko Uemura[a], Pascal Schlich[b,*]

[a] *Research Laboratories for Alcohol Beverages, Asahi Breweries, Ltd., 1-21, Midori 1-Chome, Moriya, Ibaraki 302-0106, Japan*

[b] *Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France*

*Corresponding author at: Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

E-mail address: pascal.schlich@inrae.fr (P. Schlich)

[Abstract]

While drinkability is of paramount importance when discussing beer, there is no established methodology to assess it. The main objective of the present study was to develop a new method of sensory evaluation with consumers to obtain a better understanding of beer drinkability. With the aim of being practically and effectively used in a wide range of consumer tests in an *ad libitum* consumption setting, a new method called the "Multiple-sip Drinkability Test" was developed to evaluate beer drinkability defined as "the will of drinking". The method is based on Temporal Dominance of Sensations (TDS) paired with dynamic wanting.

This paper presents the test designs and the results from two studies of four commercial Japanese beers, one conducted with an expert sensory panel and the other with a naïve panel of consumers. The results revealed the importance of monitoring dynamic wanting over sips. In both studies, product differences in *wanting*, almost nonexistent at the beginning, gradually became larger.

These studies also elicited the characteristics of beer with high drinkability. The product that attained the highest *wanting* scores in both studies was perceived as having less of a standout flavor, thereby producing fewer build-up effects on sensory perceptions, which suggests that the greater the sensory load produced by a beer, the less one wants to drink it

continuously. This methodology should be used with different types of beers and consumers to obtain a broader understanding of the sensory drivers of beer drinkability and consumer satisfaction.

[Keywords]
Temporal Dominance of Sensations (TDS), wanting, *ad libitum*, drinkability, Multiple-sip Drinkability Test, beer

[1] Introduction

1.1. Motivation for the research

It is an indubitable fact that a manufacturer's mission is to keep delivering benefits to consumers, which can be accomplished by examining the usage experience of their products. Sensory and consumer science is likely one of the best means of bridging the gap between a manufacturer and consumers caused by misunderstanding the consumers' expectations. The authors have struggled for a long time with an issue in consumer research in the beer industry; namely, why products that are liked/disliked in consumer research are not necessarily the ones that sell well/poorly in the market, and vice versa. The authors believe that the answer to this question does not simply lie in marketing investments in products, and thus began this research to search for methodologies to better understand consumer responses in a way that is relevant to consumption in real life as well as efficiently applicable to product development in the beer industry.

Liking has been used in the beer industry as a key measurement for a very long time. Liking is an easy construct to work with because the construct itself is perfectly clear and the relationship between the word label (like/liking) and the construct is unambiguous to both consumers and researchers (Thomson & Bailey, 2006). However, two beers which are liked equally at first are not always equally liked at the end of drinking. This aspect has long been discussed in the beer industry as a construct of "drinkability." In past years, the industry used consumer tests called the "Session Test" or the "Sessionability Test" in which participants were asked to drink large quantities of beer in one session and then researchers measured the total volume of consumption for each beer (Greenhoff & Buck, 2006; Dickie et al. 2006). These methods, however, include challenging issues related to ethics, health, feasibility, test controls, and so forth. Thus, the beer industry has long desired the development of a new scientific method to measure beer drinkability in consumer research.

1.2. Previous studies on beer drinkability

Several physiological studies on beer drinkability have been conducted in the last two decades:

- Nagao et al. (1998) showed a correlation between beer drinkability and gastric emptying, and suggested that measurement of the relaxed cross-sectional area of the pylorus antrum was useful to evaluate stomach fullness during beer drinking.
- They also studied the sensory perceptions of stale beer and beer with an added unpleasant taste and flavor in correlation with the urination rate, and found that the unpleasant taste and off-flavor of beer resulted in a lower urination rate while drinking beer (Nagao et al. 1999).
- Kojima et al. (2009) used a noninvasive biometric system to examine the relationship among throat sensation, beer flavor, and swallowing motion while drinking beer. They concluded that the sensation in the throat was an important factor in beer drinkability.

These types of physiological studies have disentangled some of the aspects of beer drinkability with their objective measurements. Presently, though, there is a lack of consumer-relevant subjective approaches that would add more value in understanding the construct of drinkability by linking the results with those from physiological studies. While bearing in mind that the ultimate goal of studying beer drinkability is to maximize consumers' enjoyment of drinking beer, it is of paramount importance that we understand beer drinkability from the consumers' perspective. Thus, there is a strong need to develop a consumer-relevant approach for assessing beer drinkability.

According to Mattos and Moretti (2005, 2006), drinkability can be defined as "the will of drinking or the weakness of the sensation of satiation" and "to be appreciated until the last drop." Several researchers have studied the impact of ingredients (malts, hops) and brewing technology (yeast, fermentation) on sensory perceptions linked with beer drinkability (Davies, 2006; Kaltner & Mitter, 2006; V. Opstaele et al. 2006; Taldi, 2006). Despite their efforts to clarify factors affecting beer drinkability, it should be noted that each of these researchers needed to create their own measurement as a response variable of drinkability. Unfortunately, the absence of commonly accepted methods for measuring drinkability has hampered efforts to clarify the factors affecting it. This is the impetus behind the present study aimed at developing a new method to measure "the will of drinking" with a focus on the sensory effect for beer drinkability, which could be widely used in consumer tests.

1.3. Consumer research in food and drink categories

The most popular type of consumer test with foods involves product trials at a central location. These kinds of tests offer reasonably good control conditions, and the staff can be well trained in product preparation and handling (Lawless & Heymann, 1998). Although the product is not tested under its normal conditions of use, such as at home or at parties, restaurants, etc., central location tests (CLTs) have the advantage that respondents evaluate the product under controlled conditions, which enables any misunderstandings to be cleared up (Meilgaard et al. 2016). With these points in mind, a consumer-relevant method for assessing beer drinkability that can be widely applicable in CLTs would be a valuable solution toward not only obtaining a greater scientific understanding of beer drinkability, but also for product development, particularly during the early stage when many prototypes need to be tested.

Being aware of the issue regarding CLTs, several researchers have studied the influence on product acceptance of how food and drink are consumed by respondents in CLTs. Matuszewska et al. (1997) compared three procedures for consumer assessment of fat spreads, and concluded that having consumers spread a margarine sample *ad libitum* by themselves on a standardized piece of bread was the most preferable procedure due to its superior discrimination power and labor-saving sample preparation. Zandstra et al. (1999) showed in their study on yogurts that the optimal sucrose concentration as determined by the taste-and-spit test was higher than that determined from the *ad libitum* consumption test. Posri and MacFie (2008) compared overall liking scores of tea bag products obtained from three CLTs in which the degree of freedom when making tea was different, and concluded that the provision of greater "freedom" in CLTs could result in greater discrimination between tested products. Although the influence of testing conditions on product acceptance could differ depending on the type of product, these previous studies indicated that improving the design of CLTs based on consumer insight could potentially result in a better understanding of estimating consumer response in real life.

1.4. Beer studies from the perspective of sensory and hedonic perceptions

Since the ultimate goal of studying beer drinkability is to maximize consumers' enjoyment of drinking beer, it is crucial that we investigate the factors affecting beer drinkability in terms of consumers' sensory perceptions as well as affection. Beer drinking behavior involves dynamic rather than static sensations, which implies the necessity of incorporating temporal aspects into evaluations to grasp changes in sensory perceptions over time while drinking. Pineau et al. (2003) proposed a new method of sensory evaluation, called Temporal Dominance of Sensations (TDS), in order to record several sensory attributes simultaneously over time. According to them, TDS makes it possible to collect temporal data during one single evaluation for up to 10 attributes of complex food products, and to draw curves of the dominance of each attribute over time for each product (Pineau et al. 2009). The same group of researchers also proposed an extended method of TDS, called multiple-intake (sip or bite) TDS, in order to avoid potentially misleading results from a single intake because, as Koster (2003) pointed out, the evaluation on a single intake only would not capture the changes in sensory perceptions that occur during actual consumption. This method made it possible to obtain a series of consecutive TDS evaluations at every intake, which was more relevant to consumption of a product in real life (Vandeputte et al. 2011; Pineau et al. 2013; Schlich et al. 2013). As an example, Zorn et al. (2014) applied multiple-sip TDS methodology to evaluate different sweeteners in orange juice, and showed its effectiveness in identifying differences in the dynamics of their sensory characteristics, which had not been identified using static measurements. Furthermore, Thomas et al. (2015) suggested a new methodology for characterizing temporal drivers of liking (TDL) based on the ability of consumers to record their changes in liking and to perform a TDS task.

Several researchers have recently reported on their studies of beer using temporal aspects. Vazquez-Araujo et al. (2013) compared three different techniques of temporal sensory methods for evaluating beer flavors: time intensity, TDS, and drinking profile. They demonstrated that all three techniques are useful for evaluating beer flavor, and indicated the pros and cons for each method. Simioni et al. (2018) reported a study on special beers (Bohemian Pilsner, Witbier, Belgian Strong Ale Dubbel, and Russian Imperial Stout) in which consumers took six sips (20 mL each) in total to evaluate a sample of the four different beers on four consecutive days. The tasters performed TDS, followed by a liking evaluation, for every sip of each beer sample until they finished taking the 6^{th} sip. The authors concluded that multiple-sip TDS associated with an acceptance test promoted an enhanced understanding of consumer behavior regarding temporal description and sensory acceptability over repeated consumption. Ramsey et al. (2018) applied a combination of temporal liking (TL) and temporal Check-All-That-Apply (TCATA), with the aim of determining the influence of ethanol concentration on consumer evaluation for lager beer. Participants in their study drank 30 mL of a beer sample and continued the evaluation up to 60 seconds for TL or for TCATA in a separate session, and they repeated the same evaluation method for four different beers during the same session. The researchers demonstrated the effectiveness of using these combined temporal methodologies for clarifying the influence of ethanol on the perception of liking and sensory attributes by beer consumers.

These previous studies expanded the range of approaches for sensory and consumer science in evaluating beer flavors. However, these studies were conducted using controlled settings, such as the amount of beer in each sip or the number of sips per panelist. With the caveat presented by Koster (2003) in mind, these approaches with limited and controlled consumption might not provide sufficient information to understand consumers' responses when drinking an entire beer. To the best of our knowledge, except for one study reported by Thomas et al. (2016), no other studies have investigated the dynamics of sensory and hedonic perceptions throughout the consumption of a whole package of a food or a full portion of a drink in an *ad libitum* consumption setting. Thomas et al. (2016) showed the effectiveness of using a method consisting of collecting TDS and dynamic liking data in the same session and during the consumption of the full portion of an oral nutritional supplement. However, no study on beer has been conducted from this perspective. Furthermore, when discussing topics such as beer drinkability or sense of satiation regarding food, collecting data on dynamic wanting rather than liking, as a measurement of evoked affection, would make more sense in a correlation with the volume of beer or the amount of food actually consumed.

1.5. Scope of the present study

Finally, two other topics should be mentioned when talking about beer drinkability: "context of consumption" and "emotions." Recently, Jaeger and Porcherot (2017) pointed out in their review paper that "contextual influences partially shape food-related consumer behavior," and Nijman et al. (2019) elucidated the effects of consumption context towards consumer responses to beer. These studies indicate the importance of considering consumption context when assessing beer drinkability. Moreover, several studies have recently reported the effectiveness of measuring emotions evoked when drinking beer with the aim at differentiating products beyond liking (Cardello et al. 2016; Gomez-Corona et al. 2017; Jaeger et al. 2017; Jaeger et al. 2018; Silva et al. 2019). These studies indicate the potential benefit of measuring emotions to better understand the construct of beer drinkability for consumers. These two topics, "context of consumption" and "emotions," were not directly covered in the scope of the present study since its focus was to establish a basic method to evaluate beer drinkability from a sensory perspective.

1.6. Aim of the research

The main objective of the present study was to develop a new method of sensory evaluation with consumers to obtain a better understanding of beer drinkability. The aim was to establish a new method to measure "the will of drinking", which can be practically and effectively used in a wide range of consumer tests in *ad libitum* consumption settings. The ultimate goal of this study was to clarify the factors that affect beer drinkability, which could be achieved by investigating the effects of manipulating explanatory variables such as ingredients, recipes and brewing technologies. In this way, this research could contribute to the development and marketing of products for maximizing consumers' enjoyment of drinking beer.

[2] Methods

2.1. Overview of the approach in this study

The investigation was divided into two different parts. In Phase 1, a test was conducted with an expert beer panel to confirm the capability and feasibility of our methodology. In Phase 2, a test was conducted with a consumer panel to confirm the validity of our methodology with beer consumers. This type of approach was used to acquire more implications and more aspects for discussion rather than jumping immediately into consumer tests.

2.1.1. Test design of Phase 1

In Phase 1, a test was conducted with an expert beer panel consisting of 11 employees of the R&D center at Asahi Breweries, Ltd. (Moriya, Japan) who were trained in tasting and evaluating beer. These panelists were recruited from approximately 40 panelists of the Asahi expert beer sensory panel who have from 6 months to nearly 20 years of experience in beer evaluations. These panelists evaluate beer samples regularly once or twice a day to contribute to beer sensory evaluations conducted in the R&D center. An annual training session and the daily participation in beer sensory evaluations help them maintain their ability to evaluate beer characteristics, which, in Phase 1, were based on consensus using the definition of the following sensory attributes: *malty, hoppy, estery, sweet, sour, bitter, astringent*, *and stimulating feeling.*

In Phase 1, the expert beer panel evaluated 4 products, with 2 repetitions each, for a total of 8 evaluations per panelist. Panelists in Phase 1 drank a maximum of 500 mL of each of the beer samples in 45 minutes. (See *2.1.3. Rationale behind the test designs on maximum consumption volume.*)

2.1.2. Test design of Phase 2

In Phase 2, a test was conducted with a naïve panel consisting of 44 participants (30 men and 14 women, age range: 20s to 50s) who were recruited from the Asahi Group R&D Center (Moriya, Japan). These panelists were not trained for beer evaluation, but were used to drinking beer at least once a week. A drinking frequency of at least once a week and a male-to-female ratio of 2:1 among the study group are standard recruitment criteria when conducting consumer research in the beer industry in Japan.

In Phase 2, the naïve panel evaluated 3 products, with no repetition, for a total of 3 evaluations per panelist. Panelists in Phase 2 drank a maximum of 350 mL of each sample in 30 minutes. (See *2.1.3. Rationale behind the test designs on maximum consumption volume*.)

2.1.3. Rationale behind the test designs on maximum consumption volume

In Japan, around 70% of beer sales are attributed to beers consumed at home (Fuji Keizai Co., Ltd. 2019). Most of these beers for consumption at home are sold in the regular 350 mL can or the larger 500 mL can. Since it was expected that monitoring evaluations for a longer time would result in more discrimination among products, the 500 mL can was chosen for the full portion drinkability test in Phase 1. The volume of total consumption was restricted to 500 mL to be in line with the guideline on responsible drinking (around 20 g pure alcohol per day) established by the Ministry of Health, Labour and Welfare in Japan (*Kenkou Nippon* 21).

According to a survey conducted by Asahi Breweries in 2018, among beer drinkers (2000 men and 1000 women, age range: 20s to 50s) with a drinking frequency of at least once a week, 71% of men do not consume more than 700 mL of beer (27% stop after drinking one can of 350 mL and 44% stop after drinking one can of 500 mL or two cans of 350 mL) and 92% of women do not consume more than 700 mL (56% stop after drinking one can of 350 mL and 36% stop after drinking one can of 500 mL or two cans of 350 mL) in one session at home. Thus, the investigation with beer consumers in Phase 2 focused on the size of a full portion of a regular 350 mL can, which is at least half of the total consumption volume for a majority (71% for men and 92% for women) of beer consumers in the Japanese market. Another reason for choosing the 350 mL can was its potentially high acceptability and feasibility when considering extensive use of our proposed methodology in consumer tests. If evaluations of a full portion of 350 mL lead to sufficient product discrimination, it would be considerably useful for the beer industry in product development, particularly at the early stage when many prototypes need to be tested.

2.2. Test flow

Figure 1 shows the flow of the "Multiple-sip Drinkability Test." First, the panelists take a sip and participate in TDS for 30 seconds, followed by a Liking evaluation. Second, they wait until they feel like drinking another sip. Just before they want to have the 2nd sip, they evaluate their *residual sensation*, which is the degree of flavor perceptions left in the mouth and nose. Next, they take the 2nd sip and evaluate its *ease of drinking*. Again, they wait until they feel like drinking another sip. Just before they want to have the 3rd sip, they evaluate how much they "Want to drink continuously." For the 4th, 5th, and 6th sips, they perform the same evaluation methods as those for the 1st, 2nd, and 3rd sips, respectively. Likewise, they repeat the same evaluations every 3 sips. (Note: *Residual sensation* was only asked in Phase 1. See section *2.7* for further details.)

Most importantly, both of the tests in Phase 1 and Phase 2 were conducted in an *ad libitum* setting in terms of the drinking amount and pace. Neither the amount of beer that should be drunk in each sip nor the timing of the sips was set in the tests. Panelists could also stop drinking whenever they wanted.

2.3. Test products

Four commercial beer samples marketed in Japan were ~~used~~ selected for this study. Detailed product information is not mentioned in the present paper to avoid identifying specific brands. Instead, these samples were labeled as B1, B2, B3 and B4. The main analytical and sensory characteristics of the samples are shown in Table 1. Samples B1, B3 and B4 were selected ~~as test samples in this study~~ because they were major brands sold in the market and because each possessed different sensory characteristics according to previous sensory tests evaluated by the Asahi expert beer sensory panel. Sample B2, however, was known to have quite similar characteristics to those of sample B4. Only a slight difference was previously reported between samples B4 and B2; namely, sample B2 was perceived as more hoppy and less malty than B4 in a very subtle way. The reason why we included Sample B2 was to see if our new methodology was able to detect any difference between it and B4 in the *ad libitum* setting of drinking. However, no significant difference was observed between samples B4 and B2 regarding sensory and hedonic perceptions in Phase 1 with the expert panel. Therefore, we decided to eliminate sample B2 in Phase 2 with consumers in order to simplify the test design and to reduce the respondents' burden while participating in the test.

All the test samples were served in plastic cups at a temperature of 8 (± 1) °C, which is a common temperature for evaluating beers in sensory studies in Japan. Each product was presented with its own unbranded code using a three-digit random number, according to Williams Latin Square experimental designs.

2.4. Test environments

Each panelist evaluated one beer sample per day in an individual sensory booth in the R&D center at Asahi Breweries, Ltd. During the evaluation session, the panelists were not allowed to drink water to avoid disrupting evaluations of the beer samples. They were only allowed to eat ~~non-topping salt crackers~~ plain crackers with no salt topping while they were drinking the beer samples. All the panelists, including the expert panelists in Phase 1, were instructed to drink the beer samples freely in the way they usually consumed beer. Use of phones or personal computers (PCs) for work was prohibited during the sessions, but magazines were available in the sensory booths for the participants to read. Following the instructions displayed on the PC screens in the isolated sensory booth, each panelist performed sensory and hedonic evaluations throughout the session, and the data were captured using the TimeSens© software (INRA, Dijon, France). After the panelists left the sensory room, the volume of beer actually consumed was calculated by weighing the remaining beer in the cups.

2.5. TDS evaluation

Each panelist took the 1st sip of a beer sample and participated in TDS for 30 seconds. Likewise, they repeated TDS every 3 sips, for the 4th, 7th, 10th sips, and so on. In line with Pineau et al. (2012) recommending not to include more than 10 attributes in a TDS evaluation, the TDS list in Phase 1 included eight attributes: "*malty" (roasted, caramel, grainy, worty), "hoppy" (floral, grassy, resinous, spicy, citrus), "estery" (fruity, banana, apple, soapy, solvent-like), "sweet taste" (primary taste), "sour taste" (primary taste), "bitter/astringent" (bitter as primary taste / astringent as the feeling of mouth puckering, tannin-like), "stimulating feeling" (the feeling of pricking and stinging in the mouth and throat), and "others*." These attributes were selected to sufficiently describe the sensory characteristics of the four beer samples based on previously obtained information from the Asahi expert beer sensory panel (Table 1). The expert panelists were all familiar with the definition of these attributes, and they were used to performing TDS with this list. *Others* was included as a choice in the list in case the panelists perceived some flavors that were not applicable to the seven specific attributes. In Phase 2, the naïve panelists received 5 minutes of instruction on the procedure of the TDS evaluation. The TDS list in Phase 2 included eight attributes: "*malt-like flavor," "sweet flavor," "refreshing feel," "sweet taste," "sour taste," "bitter/astringent," "stimulating feel," and "others*." These attributes had been commonly used in consumer tests with beer drinkers in Japan. The definition of each attribute was not explained to the naïve panelists as with normal consumer tests in Japan. The *others* category was also included in the attribute list in Phase 2.

During the test, the selected TDS attributes were presented on one screen, and only dominance was recorded (intensity was not recorded). The attribute order on the screen was different for each panelist, but each panelist kept his or her own order for all sips and products to avoid attribute order effects while facilitating the task of the tasters (Pineau et al. 2012, Thomas et al. 2016). They had to choose at any tasting time the dominant attribute among the eight proposed. The concept of dominant attribute was defined as "the attribute associated with the sensation catching the taster's attention at a given time." Panelists were free to choose the same attribute several times during a tasting, or never select a sensation among the eight proposed (Pineau et al. 2009, Thomas et al. 2016). They were also able to unclick the attribute buttons in case they did not feel any sensations during the 30-second period, and those moments were regarded as *No dominance* when analyzing the results.

2.6. Dynamic evaluation of wanting

Just before the panelists wanted to have the 3rd sip, they evaluated how much they "want to drink it continuously" on a linear rating scale ranging from "I do not want to drink it continuously at all" to "I want to drink it continuously very much." The panelist's indication on the linear rating scale was converted in the TimeSens© software to a score ranging from 0 to 10 for data analysis purposes (the same procedure was used for the measurements described in section 2.7.). They repeated this *wanting* evaluation every 3 sips, just before the 6th, 9th, 12th sips, and so on. Since the main focus of this study was to measure "the will of drinking" for beer drinkability, the greatest importance was placed on this dynamic evaluation of "wanting to drink continuously."

2.7. Dynamic evaluation of liking, ease of drinking and residual sensation

Immediately after the panelists completed the TDS (1st, 4th, 7th sips, and so on, every 3 sips), they indicated their *liking* for the beer sample on a linear rating scale ranging from "I do not like it at all" to "I like it very much."

Immediately after the subjects took the 2nd, 5th, 8th sips, and so on, every 3 sips, they evaluated its *ease of drinking* on a linear rating scale ranging from "not easy to drink at all" to "very easy to drink."

Subjects evaluated their *residual sensation* (from "not strong at all" to "very strong") just before they took the 2nd, 5th, 8th sips, and so on, every 3 sips only in Phase 1. This measurement was not included in Phase 2 with the naïve panel in order to simplify the test design and to reduce potential fatigue caused by evaluating many attributes. ~~with the naïve panel~~.

2.8. Data analyses

Data was analyzed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). TDS curves and TDS band plots were generated using TimeSens 1.3 (INRA, Dijon, France).

2.8.1. TDS analysis

A TDS curve represents the evolution of the dominance rate (which is the proportion of panelists who selected a certain attribute as dominant) over time. For each product, TDS curves of all the tested attributes are depicted on the same graph. The TDS curves of the differences in the dominance rate can be superimposed to compare two products. The differences are plotted only when significantly different from zero and when the higher dominant rate is significantly higher than the chance level, thereby highlighting the differences between the products over time (Pineau et al. 2009).

A TDS band plot is generated from TDS data to simplify interpretation of the TDS curves, and facilitates comparison of the sequence of the dominant attributes in all the products of the study.

2.8.2. Analysis of dynamic wanting

The analysis of variance (ANOVA) model used for analyzing the *wanting* data is:

Wanting = product + sip + panelist + product*sip + product*panelist + sip*panelist + error

In this model, the product, sip and product*sip interaction were treated as fixed effects. The panelist and its interactions with the other factors were random effects. As mentioned in section *2.2*., all the panelists drank the beer samples freely. This resulted in a different number of sips for each panelist for each sample, which requires use of a restricted maximum likelihood (REML) algorithm instead of the usual least squares algorithm. Furthermore, successive evaluations of *wanting* were not independent from sip to sip by product within each panelist. Therefore, a repeated mixed model ANOVA was used with a first-order heterogeneous auto-regressive covariance structure among sips within panelist by product. This structure assumes that the covariance between two *wanting* scores from two sips of the same evaluation decreases with the number of sips in between these two sips.

The least square mean (LS-mean) and standard error were subsequently computed for each product, for each sip, and for each sip by each product, which enabled interpretation of the statistical differences among products, among sips, and among sips by each product, respectively. This analysis was accomplished using PROC MIXED in SAS 9.4.

2.8.3. Analysis of dynamic liking, ease of drinking and residual sensation

The same repeated mixed model ANOVA was used for other dynamic measurements, which included *liking, ease of drinking* and *residual sensation.*

2.8.4. Evolution of attribute dominance durations

Duration of dominance is obtained at the individual level by adding all the time periods during which the attribute was cited as dominant, regardless of the moment of perception (Galmarini et al. 2017). Data for the dominance duration of each attribute were analyzed by the same repeated mixed model ANOVA as those used for *wanting* and the other measurements.

[3] Results

3.1. Phase 1 results with the expert panel

3.1.1. Dynamic evaluation of wanting, liking, ease of drinking and residual sensation

Table 2a and Table 2b show results of the ANOVA with the expert panel. Fixed effects of both Sip and Product were observed for *wanting*, *liking* and *ease of drinking* at a statistically significant level ($p < 0.05$) (Table 2a), but not for *residual sensation*. Sip effect here refers to changes in evaluations over sips while drinking. The result also showed a significant Product by Sip interaction for *wanting* ($p < 0.05$). The p-values of Sip effect by Product (Table 2b) indicate that *wanting* for samples B1 and B3 significantly changed (decreased) while drinking, whereas *wanting* for samples B2 and B4 did not. Interestingly, the product differences in *wanting*, which were almost nonexistent at the beginning of Sip 3 (Sip Cycle 1), gradually became larger while drinking (Figure 2). As a result, the LS-means of *wanting* show a significant decrease in samples B1 ($p = 0.0000$) and B3 ($p = 0.0000$), but not in samples B2 ($p = 0.2367$) and B4 ($p = 0.1576$) (Table 2b). In this way, we found that samples B2 and B4 had higher beer drinkability than samples B1 and B3.

Figure 3 shows the evolution of *ease of drinking*. In contrast to the evolution of *wanting* (Figure 2), the product differences were already existent at the beginning of Sip 2 (Sip Cycle 1). In particular, this tendency was clearer when looking at the results of sample B1 compared with those of sample B4. *Wanting* for sample B1 was comparable to that for sample B4 at Sip 3 (Sip Cycle 1); however, *ease of drinking* for sample B1 was clearly lower than that for sample B4 at Sip 2 (Sip Cycle 1). The products with lower *wanting* at the end (samples B1 and B3) were lower in *ease of drinking* from the beginning. These results indicate that *ease of drinking* at each sip could lead to "wanting to drink continuously," that is, beer drinkability.

The measurements of *wanting* and *liking* were similar in that both of them detected a significant Sip effect ($p = 0.0000$ for *wanting*, $p = 0.0023$ for *liking*) and Product effect ($p = 0.0037$ for *wanting*, $p = 0.0155$ for *liking*) (Table 2a). Indeed, the correlation coefficient between *wanting* and *liking* was high at any sip cycle until Sip Cycle 6, which was 0.79, 0.87, 0.82, 0.86, 0.82, 0.83, respectively computed over 2 replicates, 11 panelists and 4 products ($n = 88$). However, as shown in Figure 4, in which the LS-mean scores over products in

Table 2b were plotted in order of sip, decreases in evaluation scores became more prominent for *wanting* than for *liking* after the middle of the drinking session (Sip Cycle 4). Also, product by sip interaction was only significant for *wanting* ($p = 0.0310$), but not for *liking* ($p = 0.2672$) (Table 2a). These findings indicate that asking about *wanting* could be more powerful in understanding beer drinkability as well as discriminating beer products, rather than asking about *liking*.

Although we found that samples B2 and B4 were significantly higher than samples B1 and B3 for *wanting* and *ease of drinking* ($p < 0.05$ for both), we could not detect any differences between samples B2 and B4 or between samples B1 and B3 for these measurements (Table 2a).

3.1.2. Dynamic evaluation (TDS) of flavor characteristics

Figure 5 shows the evolution of sensory perceptions. The TDS band plot is a summary of the TDS results for the four beer samples over time while drinking. From the left side, the TDS results are shown every 30 seconds in order of time for the 1st, 4th, 7th sip, and so on. The colors indicate the sensory characteristics that were significantly dominant during each second. For sample B1, "*Hoppy*" (blue) was perceived to be consistently dominant at the beginning of each sip and "*Bitter/Astringent*" (red) became more dominant over time while drinking. For sample B3, "*Bitter/Astringent*" was consistently dominant for most of the time from beginning to the end. Samples B2 and B4 were perceived to be quite similar to each other, with "*Malty*" (purple) and "*No Dominance*" (black) more dominant than the other two products. As previously mentioned in section *2.5. TDS evaluation,* the moment when a panelist did not feel any sensations was regarded as "*No Dominance*" when analyzing the results. In summary, when comparing the products' characteristics, sample B1 was characterized as the most *Hoppy*, sample B3 as the most *Bitter/Astringent*, and samples B2 and B4 had the longest period of "*No Dominance*", which can be seen until the end of drinking.

3.1.3. Link between dynamic wanting and sensory perceptions

The relationship between the results of dynamic *wanting* and TDS for the four tested products reveals some indications for beer drinkability and sensory perceptions. The products that scored higher in *wanting*, samples B2 and B4 (Table 2a and Figure 2), were perceived as having less standout flavors such as "*Hoppy*" and "*Bitter/Astringent*" (Figure 5), thereby producing less build-up effects on sensory perceptions. These results indicated that the greater the sensory load produced by a beer, the less one wants to drink it continuously.

3.2. Phase 2 results with the naïve panel

3.2.1. Dynamic evaluation of wanting, liking and ease of drinking

Table 3a and Table 3b show the results of the ANOVA with the naïve panel. Fixed effects of both Sip and Product were observed for *wanting*, *liking* and *ease of drinking* at a statistically significant level ($p < 0.05$). Again, Sip effect means that the evaluation scores change over sips while drinking. The p-value for product by sip interaction for *wanting* ($p = 0.1767$) was relatively lower, but not at a statistically significant level in the naïve panel. Similar to the results from Phase 1 with the expert panel, the p-values of Sip effect by Product indicate that *wanting* for samples B1 ($p = 0.0101$) and B3 ($p = 0.0001$) significantly changed (decreased) while drinking, but not for sample B4 ($p = 0.1893$) (Table 3b). Notably, the product differences in *wanting*, which were almost nonexistent at the beginning, gradually became larger while drinking (Figure 6). This finding is consistent with the results of Phase 1 with the expert panelists. The LS-mean of *wanting* for sample B4 was significantly higher than that for sample B3 ($p < 0.0001$), and ~~almost significantly~~ tended to be higher than that for sample B1 ($p = 0.1007$) (Table 3a).

For the results with the naïve panel, *wanting* and *liking* were similar in that both of them detected a Sip effect ($p < 0.0001$ for *wanting*, $p = 0.0040$ for *liking*) and Product effect ($p < 0.0001$ for both *wanting* and *liking*) and differences in product LS-means ($p < 0.05$) at statistically significant levels (Table 3a). The correlation coefficients between *wanting* and *liking* were high throughout all of the sip cycles, which were 0.77, 0.85, 0.87, 0.89, 0.94, 0.89 and 0.89, computed over panelist by product ($n = 126$). As shown in Figure 7, in which the LS-mean scores over products in Table 3b are plotted in the order of sip, development of decreases in these scores seem to be in alignment with each other, and no strong evidence was observed in the naïve panel regarding whether asking about *wanting* was more powerful than asking about *liking*. Compared with the expert panelists, the naïve panelists tended to use a narrow range on the linear rating scale while giving higher *wanting* scores overall, and they were less inclined to give lower scores during the entire experiment.

3.2.2. Dynamic evaluation (TDS) of flavor characteristics

Figure 8a shows TDS difference curves between samples B4 and B3, which were respectively the highest and the lowest in *wanting* scores (Table 3a). The sensory attributes over the centerline indicate a significantly higher ($p < 0.1$) dominance rate for sample B4, and those below the centerline indicate a significantly higher ($p < 0.1$) dominance rate for sample B3 over the 30-second period, within each sip. (Note: In Figure 8a and Figure 8b, "within sip" analyses were used to focus on direct comparison of sensory perceptions between two products.) Sample B4 scored significantly higher for "*Refreshing Feel*" in the beginning and for "*No Dominance*" at the end. In contrast, sample B3 scored significantly higher for "*Bitter/Astringent*" from the beginning to the end.

Figure 8b shows TDS difference curves between samples B4 and B1, which were respectively the highest and the second highest in *wanting* scores (Table 3a). More dominance could be seen for sample B4 in "*Stimulating Feeling*" and "*Malt-like Flavor*." However, the difference between samples B4 and B1 was not as clear as that observed by the expert panel, who perceived sample B1 as being more "*Hoppy*."

To grasp a better understanding of the panelists' responses, we conducted a segmentation analysis for samples B4 and B1 (Figure 9). The left graph shows TDS difference curves among the panelists, who gave a higher *wanting* score for sample B4. Conversely, the right graph shows curves among those that gave a higher *wanting* score for sample B1. Interestingly, perceptional differences were found between these two groups. The panelists who scored sample B4 higher in *wanting* perceived it as less dominant for "*Bitter/Astringent*," whereas it was the opposite for those who scored sample B1 higher for *wanting*. Because of the small number of panelists ($n = 42$) in this study, these findings from the segmentation analysis should be examined with caution and need to be validated with a larger number of consumers.

3.2.3. Development of bitter/astringent dominance duration

Figure 10 shows the development of dominance duration for "*Bitter/Astringent*" for all three products. Dominance duration refers to the total number of seconds that "*Bitter/Astringent*" was selected as dominant in each sip. Sample B3 was significantly higher ($p < 0.01$) for the dominance duration of "*Bitter/Astringent*" compared to samples B1 and B4, and B3 showed a tendency to accumulate "*Bitter/Astringent*" over time during the drinking session. Considering the fact that sample B3 was lower in *wanting* compared to samples B1 and B4, it can be assumed that the build-up of "*Bitter/Astringent*" could lead to low drinkability.

3.3. Wanting and volume of beer actually consumed

Figure 11a shows the volume of beer actually consumed by the expert panelists in Phase 1. Samples B2 and B4 tended to be consumed more than samples B1 and B3, but not at a statistically significant level because of the large variance observed among the panelists. By contrast, a significant difference in *wanting* scores was observed between the products ($p < 0.05$) (Table 2a). Therefore, in this case, asking about *wanting* was more useful for discriminating products than collecting data on the total volume consumed.

Similar results were found in Phase 2 with the naïve panelists (Figure 11b). No significant differences in consumed volume of beer were observed, while the trend in consumed volume was in line with the *wanting* scores (Table 3a). However, significant differences ($p < 0.05$) were found in *wanting* scores between products (Table 3a), with an indication that asking about *wanting* outperformed measuring the total volume of consumption in terms of discriminating products.

[4] Discussion

The present work aimed at developing a new method of sensory evaluation for beer drinkability, which can be widely and easily applied in consumer tests. On the construct of beer drinkability defined as "the will of drinking," a new method called the "Multiple-sip Drinkability Test" was developed based on TDS paired with dynamic wanting to drink continuously. In the present study, with no lengthy training, both the expert panelists and the naïve panelists were able to conduct TDS repeatedly while drinking a full portion of beer up to 500 mL or 350 mL, respectively. The panelists were able to understand TDS and implement it properly after receiving only 5 minutes of instruction. This is in line with the suggestions of Monaco et al. (2014) and Schlich (2017). Furthermore, the participants in this study only needed to follow the instructions displayed on PC screens to perform their evaluations readily and intuitively on a web site specifically allocated for each respondent. These aspects of the present study will contribute as part of a widely applicable approach to sensory and consumer research for beer drinkability in CLTs.

To develop new, successful products, companies should gain a deep understanding of "the voice of the consumer" in the early stage of product development; however, this step is often ignored or poorly executed (van Kleef et al. 2005). Koster (2003) raised issues concerning researchers' bias of subject consistency and possibly misleading outputs from consumer studies that rely only on single hedonic evaluation. The results from the present study were in agreement with Koster's statement. By measuring dynamic *wanting* while drinking a full portion of beer, product differences in *wanting*, which were almost nonexistent at the beginning, were found to gradually become larger while drinking. Investigations were also conducted on the panelists' responses for both *wanting* and *liking* asked dynamically over sips to find a relationship between these two evaluation scales. Phase 1 results with the expert beer panel showed a superiority of measuring *wanting* over *liking* in terms of identifying product differences over time. In contrast, no large differences were observed between these two evaluations in Phase 2 with the naïve panel, who used a relatively narrow range on the linear rating scale. We restricted the maximum volume of consumption to 350 mL in Phase 2 with the naïve panel (see section *2.1.3. Rationale behind the test* *designs on maximum consumption volume*); however, increasing the volume of maximum consumption up to 500 mL might lead consumer panelists to use a broader range of the linear rating scale for *wanting*, which could result in better discrimination of beers at the end of drinking. Furthermore, when considering that some consumer groups, such as heavy drinkers are likely to express their "will" of drinking beer continuously in a clearer attitude than to show their appreciation toward a beer, *wanting* would probably be a more valuable evaluation scale if a researcher needs to choose one of these two scales for studying beer drinkability.

The test results also indicated that *ease of drinking* at each sip could lead to "wanting to drink continuously," in other words, beer drinkability. This finding is useful from the perspective of product development, particularly during the early stage when many prototypes need to be tested. By using this knowledge, asking about the *ease of drinking* for just a portion of a beer sample, rather than the full volume every time, could work as a good predictor of beer drinkability for the purpose of screening prototypes, allowing the researcher to move on to further steps with a decreased number of candidates.

Since the focus in the present work was mostly methodological and not to investigate details of beer flavors, we conducted the tests with a relatively small number of panelists. However, in addition to the development of the methodology, which enabled the discrimination of beers in terms of drinkability, the conducted study also elicited the flavor characteristics of beer with high drinkability. The results indicated that a product with less standout flavor, thereby producing fewer build-up effects on sensory perceptions, could be a beer with higher drinkability. In other words, the greater the sensory load produced by a beer, the less one wants to drink it continuously. These findings are in agreement with the results of some previous beer studies (Guinard et al. 1998; Mattos & Moretti, 2005, 2006; Davies, 2006; Greenhoff & Buck, 2006; Kojima et al. 2009).

The ultimate goal for the investigation of beer drinkability was to maximize consumers' enjoyment of drinking beer. Since the present evaluation included only four commercial products marketed in Japan and a limited number of panelists, more tests with designed prototypes sampled by general consumers on a larger scale are necessary in order to clarify further the factors affecting beer drinkability. Since the effectiveness of using the "Multiple-sip Drinkability Test" was validated with consumers, it will be possible to further investigate the impacts of various factors such as ingredients, recipes and brewing technologies by manipulating those variables when making prototypes for consumer tests. This study, particularly in Phase 2 with consumers, also showed the possibilities of perceptional differences and directional differences for highly drinkable beers depending on consumer segments. Applying this method with a large number of target consumers would be beneficial for marketing and product development in the beer industry.

Results from the tests with the expert and the naïve panels showed that asking about *wanting* was more useful for discriminating products than collecting data on the total volume consumed. One of the reasons that no significant difference in the total consumption volume was observed among the products was because of the large variance observed among the panelists. Another reason was because of the small variance in the total consumption among the products for each panelist. Possibly, with a limited amount of consumption up to 500 mL or 350 mL, "the will of drinking" can be more readily captured by simply asking about *wanting* at a conscious level rather than letting a panelist show it by drinking behavior at a subconscious level. Since the maximum amount of a beer sample was the normal volume consumed daily by a panelist, there might have been psychological barriers to stop drinking a test sample unless its flavors strongly discouraged a panelist's will of drinking continuously. This may also suggest that there could be inertia in beer drinking behavior; that is, beer could be a type of drink that is consumed continuously without a specific strong feeling of *wanting* while drinking. However, the reason for this finding remains unclear and warrants further investigation.

In the present study, the panelists were instructed to only eat ~~non-topping salt crackers~~ plain crackers with no salt topping while they were drinking the beer samples. However, considering beer drinking contexts in real life, further investigations of beer drinkability "with food" would add more value for product development and marketing purposes in the future. King et al. (2004) reported that overall liking of a flavored iced tea significantly increased when it was served together with a salad and a pizza compared with the overall liking of the tea alone. Galmarini et al. (2016) conducted a study involving multi-intake TDS to evaluate the influence of cheese on wine perception, and the results of their study showed that the dominance duration of astringency and sourness for a red wine was reduced by cheese consumption between sips of the wine. These previous investigations suggest an effectiveness of taking into consideration the possibility of "meal context" affecting beer drinkability. With the effectiveness of using the "Multiple-sip Drinkability Test" with consumers in an *ad libitum* consumption setting demonstrated in the present study, it would be worth expanding this protocol to search for the optimal pairing with foods in the construct of beer drinkability for consumers.

[5] Limitations and further research

According to Mattos and Moretti (2005, 2006), there are four different effects of beer drinkability: sensory, cognitive, post-ingestive and post-absorptive. Among these, our focus was on the sensory aspect of drinkability. For measuring it at a conscious level, the quantities of consumption in the present study were limited to 500 mL. Consequently, it would not be possible to discuss the results from the aspects of post-ingestive or post-absorptive, which would require participants to drink large quantities of beer in one session. The consumer panelists in Phase 2 drank a maximum of 350 mL of each sample, which is the volume of a full portion of a regular can available in Japan. Therefore, the present method with consumers has a limitation that it could at most estimate the power to evoke *wanting* for the second 350 mL can. As mentioned in section ~~*"2.1.1. Test design of Phase 2,"*~~ *2.1.3. Rationale behind the test designs on maximum consumption volume,* however, the majority of beer consumers in Japan do not consume more than two cans of beer in one session at home. This is why the authors believe that this method is a consumer-relevant approach. In terms of product discrimination, increasing the maximum consumption volume up to 500 mL in future studies could result in better discrimination of beers.

In the present study, each panelist conducted their evaluation of beer individually in an isolated sensory booth. They were allowed to read magazines as they wanted during a test so that they could feel at ease; however, the test environment was not ideally designed to be or to make them feel as if they were in a "real" consumption context. As recently reported (Jaeger & Porcherot, 2017; Nijman et al. 2019), consumption context could be one of the key factors affecting beer drinkability. The present method was useful in assessing beer drinkability with consumers, so it would be valuable to evaluate beers in some different consumption contexts or evoked contexts by manipulating test conditions in future studies. It should also be mentioned that the test samples were served in plastic cups, although consumers usually drink beer in beer glasses or directly from cans in real life. This might have an effect on perceived sensation, possibly caused by mouthfeel when taking a sip, amount of beer taken in each sip, changes in beer temperature while drinking, and so forth. Depending on the study objectives, this aspect should be taken into consideration when designing tests for future research.

As mentioned previously, the focus in this study was on the sensory aspect of beer drinkability. Thus, no list was provided for the panelists to assess their emotions while drinking beer samples. Given that the emotional aspect was proven to be effective in differentiating beers beyond liking (Cardello et al. 2016; Gomez-Corona et al. 2017; Jaeger et al. 2017; Jaeger et al. 2018; Silva et al. 2019), the authors are considering further studies that incorporate measurements of emotions into the current protocol, such as a combination of TDS and Temporal Dominance of Emotions (Jager et al. 2014) paired with dynamic wanting to better understand the construct of beer drinkability for consumers.

[6] Conclusion

A new method called the "Multiple-sip Drinkability Test" was developed to evaluate beer drinkability defined as "the will of drinking," based on Temporal Dominance of Sensations paired with dynamic wanting. To the best of our knowledge, this is the first published methodology with consumers in an *ad libitum* setting, which could be widely and easily applied in central location tests to obtain a better understanding of beer drinkability.

The present method enabled the discrimination of beers in terms of drinkability, and the results revealed the importance of monitoring dynamic wanting over sips. It was found that product differences in wanting, which were almost nonexistent at the beginning, gradually became larger while drinking. The results of this study also indicated that a product with less standout flavor, thereby producing fewer build-up effects on sensory perceptions, could be a beer with higher drinkability. In other words, the greater the sensory load produced by a beer, the less one wants to drink it continuously.

Since the present work was mostly methodological with a relatively small number of panelists, more work will be required to clarify further the factors that affect beer drinkability for target consumers on a larger scale. This could be achieved by testing designed prototypes with this evaluation method to investigate the impacts of changing variables, such as, ingredients, recipes and brewing technologies.

[Acknowledgements]

We thank all the volunteer panelists who participated in our tests.
This research was funded by Asahi Breweries, Ltd.
We would also like to thank the reviewers who contributed to improve significantly the quality of the manuscript.

[References]

Cardello, AV., Pineau B., Paisley, AG., Roigard, CM., Chheang, SL., Guo, LF., Hedderley DI., & Jaeger, SR. (2016). Cognitive and emotional differentiators for beer: An exploratory study focusing on "uniqueness." *Food Quality and Preference, 54,* 23–38.

Davies, N. (2006). Malt - a vital part of the brewer's palette. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Di Monaco, R., Su, C., Masi, P., & Cavella, S. (2014). Temporal dominance of sensations: A review. *Trends in Food Science & Technology, 38,* 104–112.

Dickie, KH., Parker, DK., & Muller, RE. (2006). Fancy another beer - same again? *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Fuji Keizai Co., Ltd. (2019). *Foodstuff Marketing Handbook 2020: Vol. 2.*

Galmarini, MV., Loiseau, AL., Visalli, M., & Schlich, P. (2016). Use of multi-intake temporal dominance of sensations (TDS) to evaluate the influence of cheese on wine perception. *Journal of Food Science, 81,* S2566–S2577.

Galmarini, MV., Visalli, M., & Schlich, P. (2017). Advances in representation and analysis of mono and multi-intake Temporal Dominance of Sensations data. *Food Quality and Preference, 56,* 247–255.

Greenhoff, K. & Buck, DT. (2006). Consumer measurement of drinkability. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Gomez-Corona, C., Chollet, S., Escalona-Buendia HB., & Valentin, D. (2017). Measuring the drinking experience of beer in real context situations. The impact of affects, senses, and cognition. *Food Quality and Preference, 60,* 113–122.

Guinard, J., Souchard, A., Picot, M., Rogeaux, M., & Sieffermann, J. (1998). Sensory determinants of the thirst-quenching character of beer. *Appetite, 31,* 101-115.

Jaeger, SR., Cardello, AV., Chheang SL., Beresford, MK., Hedderley, DI. & Pineau, B. (2017). Holistic and consumer-centric assessment of beer: A multi-measurement approach. *Food Research International, 99,* 287–297.

Jaeger, SR., Xia, Y., Le Blond, M., Beresford, MK., Hedderley, DI., & Cardello, AV. (2018). Supplementing hedonic and sensory consumer research on beer with cognitive and emotional measures, and additional insights via consumer segmentation. *Food Quality* *and Preference, 73,* 117–134.

Jaeger, SR. & Porcherot, C. (2017). Consumption context in consumer research: methodological perspectives. *Current Opinion in Food Science, 15,* 30–37.

Jager, G., Schlich, P., Tijssen, I., Yao, J., Visalli, M., De Graaf, C., & Stieger, M. (2014). Temporal dominance of emotions: Measuring dynamics of food-related emotions during consumption. *Food Quality and Preference, 37,* 87–99.

Kaltner, D. & Mitter, W. (2006). Possible impact of hops on taste and drinkability of beer. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Kenkou Nippon 21 (Health Japan 21). Ministry of Health, Labour and Welfare.

King, SC., Weber, AJ., Meiselman, HL., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference, 15,* 645–653.

Kojima, H., Kaneda, H., Watari, J., Nakamura, Y., & Hayashi, T. (2009). Relationships among throat sensation, beer flavor, and swallowing motion while drinking beer. *J. Am. Soc. Brew. Chem. 67(4),* 217-221.

Köster, E. P. (2003). The psychology of food choice: Some often encountered fallacies. *Food Quality and Preference, 14,* 359–373.

Lawless, HT., & Heymann, H. (1998). *Sensory Evaluation of Food: Principles and Practices, Chapman & Hall, New York*.

Mattos, R. (2006). Recent review on beer drinkability. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Mattos, R., & Moretti, R. (2005). Beer drinkability - A review. *Tech. Q. Master Brew. Assoc. Am. 42,* 13-15.

Matuszewska, I., Barylko-Pikielna, N., Szczecinska, A., & Radzanowska, J. (1997). Comparison of three procedures for consumer assessment of fat spreads: Short report. *Polish Journal of Food and Nutrition Sciences, 6,* 139–142.

Meilgaard, M., Civille, GV. & Carr, BT. (2016). *Sensory Evaluation Techniques: 5th Edition, CRC Press, New York*.

Nagao, Y., Kodama, H., Yamaguchi, T., Yonezawa, T., Taguchi, A., Fujino, S., Morimoto, K., & Fushiki, T. (1999). Reduced urination rate while drinking beer with an unpleasant taste and off-flavor. *Biosci. Biotechnol. Biochem. 63,* 468-473.

Nagao, Y., Kodama, H., Yonezawa, T., Taguchi, A., Fujino, S., Nakahara, K., Haruma, K., & Fushiki, T. (1998). Correlation between the drinkability of beer and gastric emptying. *Biosci. Biotechnol. Biochem. 62,* 846-851.

Nijman, M., James, S., Dehrmann F., Smart, K., Ford, R. & Fort, J. (2019). The effect of consumption context on consumer hedonics, emotional response and beer choice. *Food Quality and Preference, 74,* 59–71.

Pineau, N. (2013). Extension of TDS to multi-bite evaluation. In *10th Pangborn sensory science symposium, 11–15th August 2013, Rio de Janeiro, Brazil.*

Pineau, N., Cordelle, S., & Schlich, P. (2003). Temporal dominance of sensations: A new technique to record several sensory attributes simultaneously over time. In *5th Pangborn, symposium, July 20–24 (pp. 121), Boston.*

Pineau, N., Goupil de Bouillé, A., Lepage, M., Lenfant, F., Schlich, P., & Martin, N., et al. (2012). Temporal Dominance of Sensations: What is a good attribute list? *Food Quality and Preference, 26(2),* 159–165.

Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., & Imbert, A., et al. (2009). Temporal Dominance of Sensations: Construction of the TDS curves and comparison with time–intensity. *Food Quality and Preference, 20(6),* 450–455.

Posri, W., & MacFie, H. (2008). The influence of testing context on tea bag product acceptance in central location tests. *Journal of Sensory Studies, 23,* 835–851.

Ramsey, I., Ross, C., Ford, R., Fisk, I., Yang, Q., Gomez-Lopez, J., & Hort, J. (2018). Using a combined temporal approach to evaluate the influence of ethanol concentration on liking and sensory attributes of lager beer. *Food Quality and Preference, 68,* 292–303.

Schlich, P. (2017). Temporal dominance of sensations (TDS): A new deal for temporal sensory analysis. *Current Opinion in Food Science, 15,* 38–42.

Schlich, P., Urbano, C., & Visalli, M. (2013). Multi-bites or multi-sips TDS with untrained subjects: A live demonstration on chocolates. In *10th Pangborn sensory science symposium, 11–15th August 2013, Rio de Janeiro, Brazil.*

Silva, AP., Voss, H-P., van Zyl H., Hogg, T., de Graaf, C., Pintado, M. & Jager, G. (2019). Effect of adding hop aroma in beer analysed by temporal dominance of sensations and emotions coupled with temporal liking. *Food Quality and Preference, 75,* 54–63.

Simioni, SCC., Ribeiro, MN., de Souza, VR., Nunes, CA., & Pinheiro, ACM. (2018). Multiple-sip temporal dominance of sensations associated with acceptance test: a study on special beers. *J Food Sci Technol, 55(3),* 1164–1174.

Taldi, B. (2006). Yeast and fermentation, tools for increasing drinkability. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Thomas, A., van der Stelt, AJ., Prokop, J., Lawlor, JB., & Schlich, P. (2016). Alternating temporal dominance of sensations and liking scales during the intake of a full portion of an oral nutritional supplement. *Food Quality and Preference, 53,* 159–167.

Thomas, A., Visalli, M., Cordelle, S., & Schlich, P. (2015). Temporal drivers of liking. *Food Quality and Preference, 40,* 365–375.

Thomson, D. & Bailey, P. (2006). Psychology of drinkability. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Vandeputte, A., Romans, J., Pineau, N., & Lenfant, F. (2011). Innovative methods to assess the evolution of the sensory characteristics during the tasting of a full product portion (several bites). In *9th Pangborn sensory science symposium, 4–8th September 2011, Toronto, Canada.*

van Kleef, E., van Trijp, HCM., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preference, 16,* 181–201.

van Opstaele, F., Goiris, K., Syryn, E., de Cooman, L., & Aerts, G. (2006). Novel hop-derived ingredients and drinkability. *European Brewery Convention, Monograph 34, Symposium, Drinkability, Edinburgh, United Kingdom, November 2006.*

Vazquez-Araujo, L., Parker, D., & Woods, E. (2013). Comparison of temporal-sensory methods for beer flavor evaluation. *Journal of Sensory Studies, 28,* 387–395.

Zandstra, EH., de Graaf, C., van Trijp, HCM., & van Staveren, WA. (1999). Laboratory hedonic ratings as predictors of consumption. *Food Quality and Preference, 10,* 411–418.

Zorn, S., Alcaire, F., Vidal, L., Giménez, A., & Ares, G. (2014). Application of multiple-sip temporal dominance of sensations to the evaluation of sweeteners. *Food Quality and Preference, 36,* 135–143.

Figure 1. Test flow of "Multiple-sip Drinkability Test."

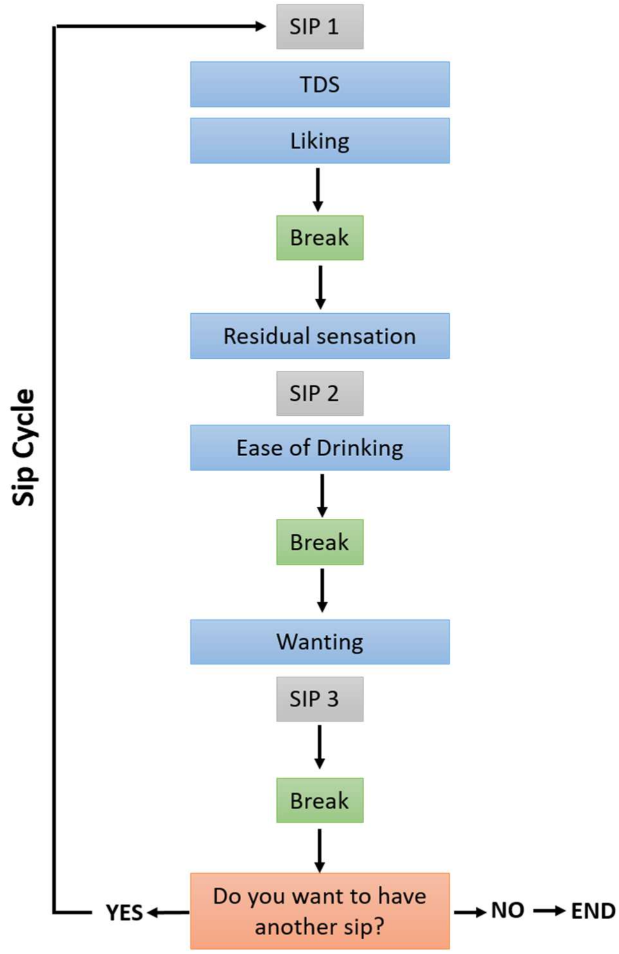

Figure 2. Development of *wanting* - Expert Panel

Product by Sip (Means ± Standard Error x 2; replicates pooled).

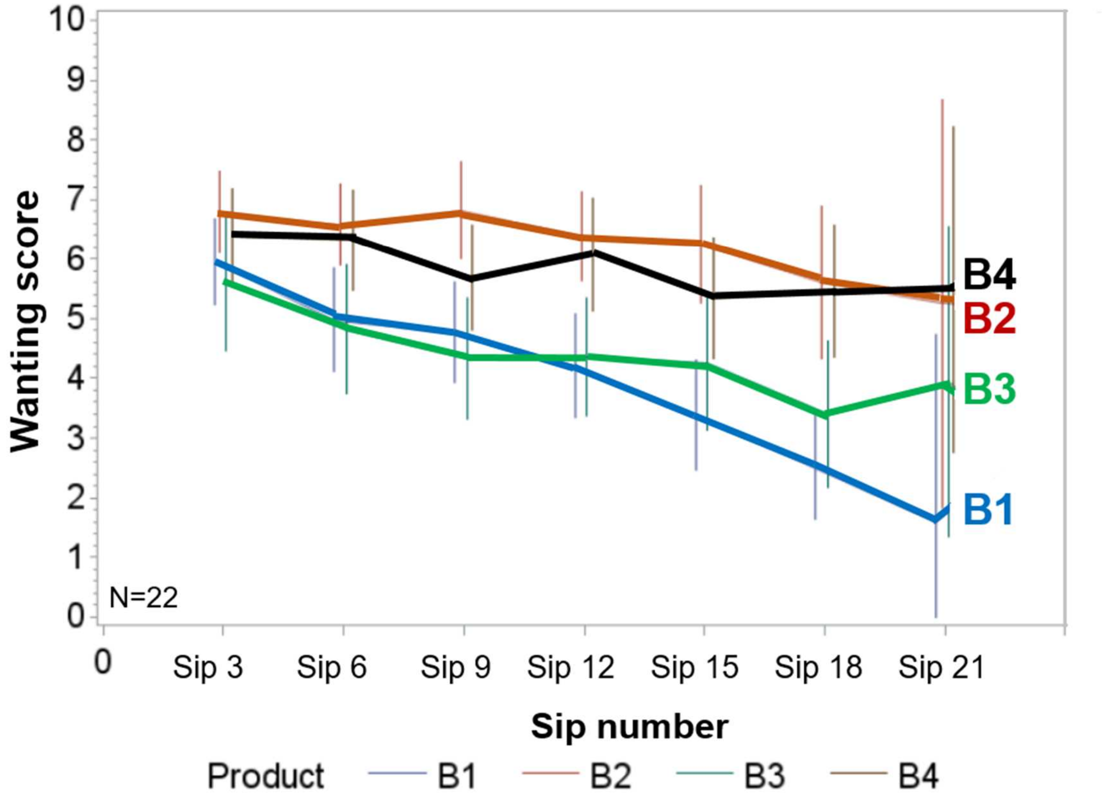

Figure 3. Development of *ease of drinking* - Expert Panel

Product by Sip (Means ± Standard Error x 2; replicates pooled).

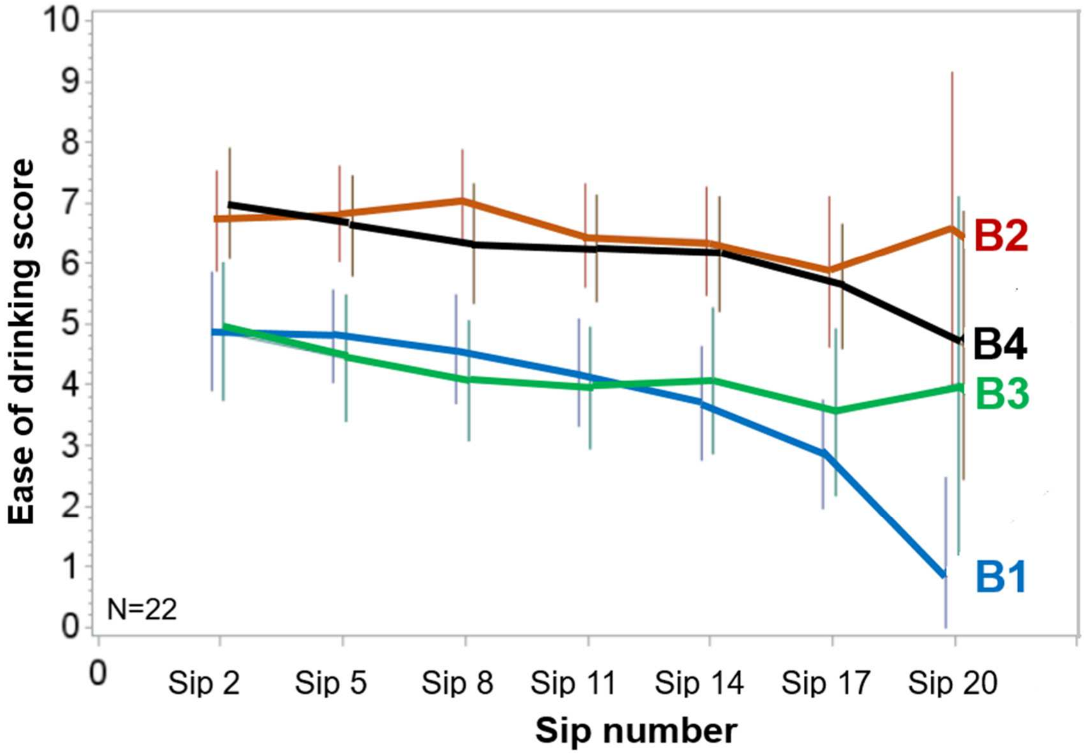

Figure 4. Development of *wanting* compared with *liking*: Sip Cycle LS-Means over products - Expert Panel.

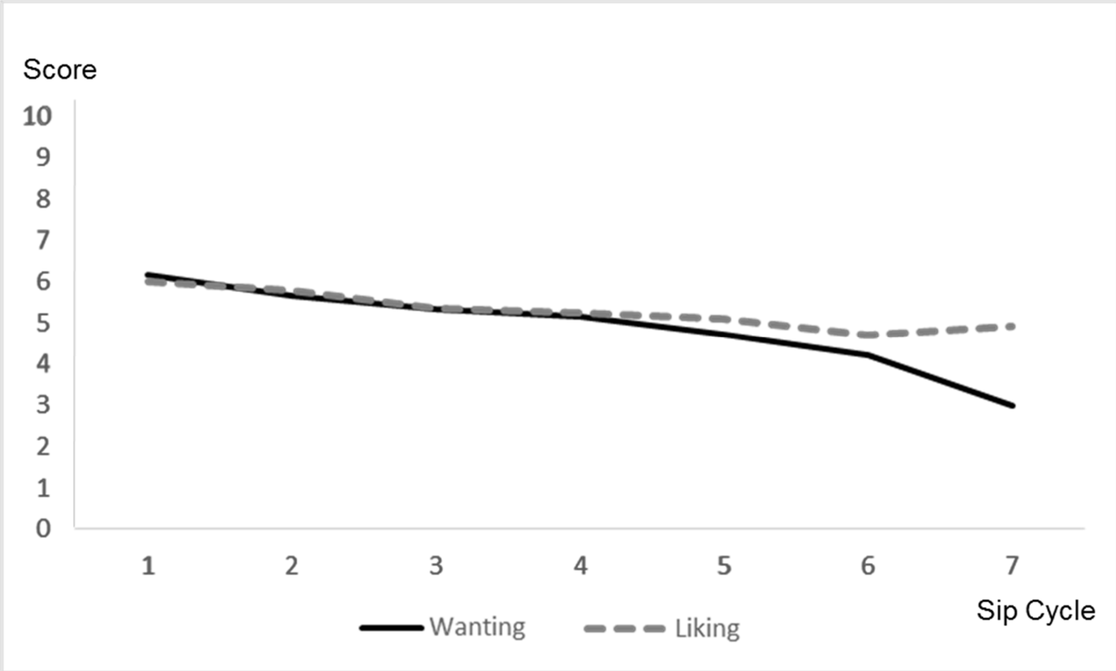

Figure 5. Development of sensory perceptions: TDS Band Plot and Products' characteristics.

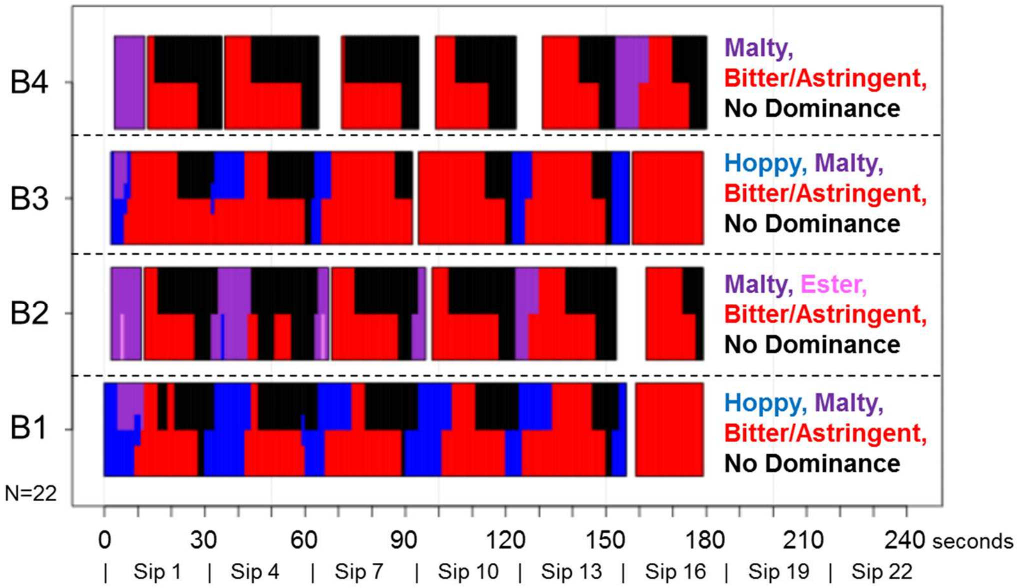

Figure 6. Development of *wanting* - Naïve Panel
Product by Sip (Means ± Standard Error x 2).

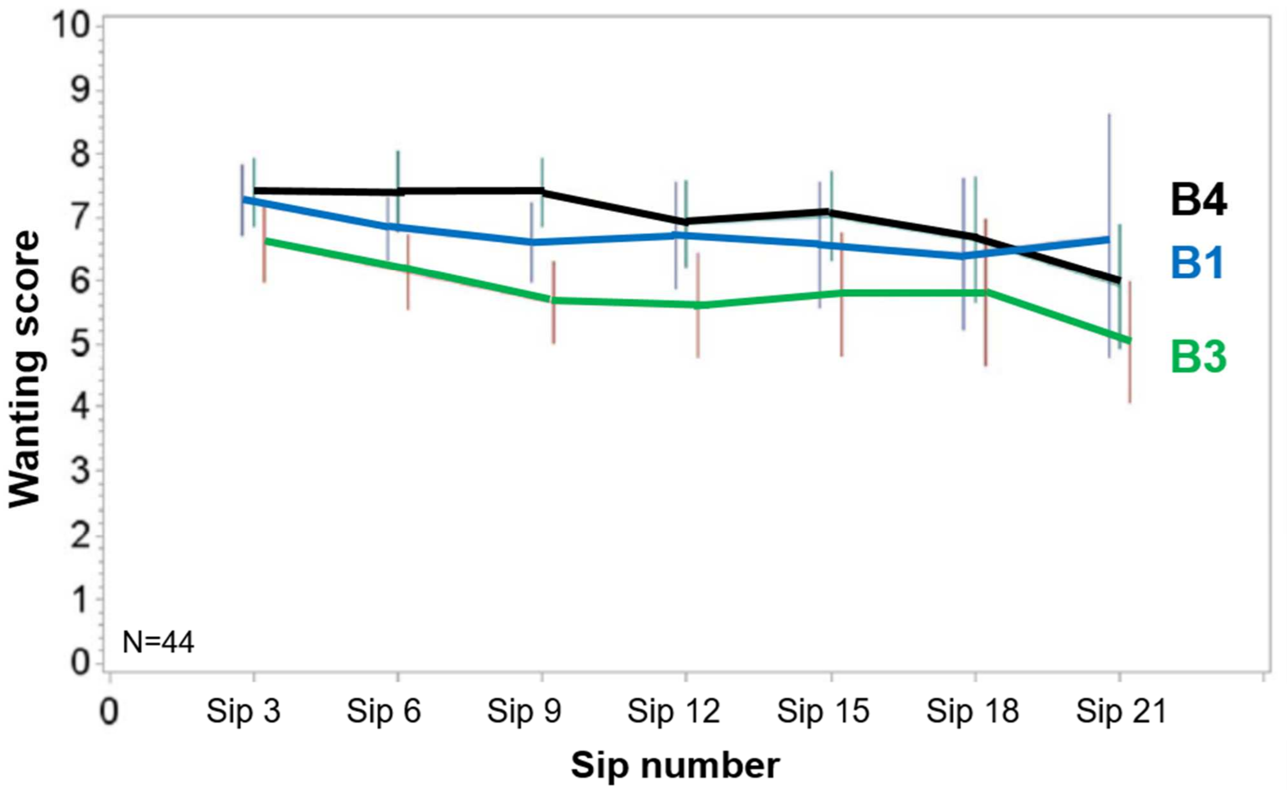

Figure 7. Development of *wanting* compared with *liking*: Sip Cycle LS-Means over products - Naïve Panel.

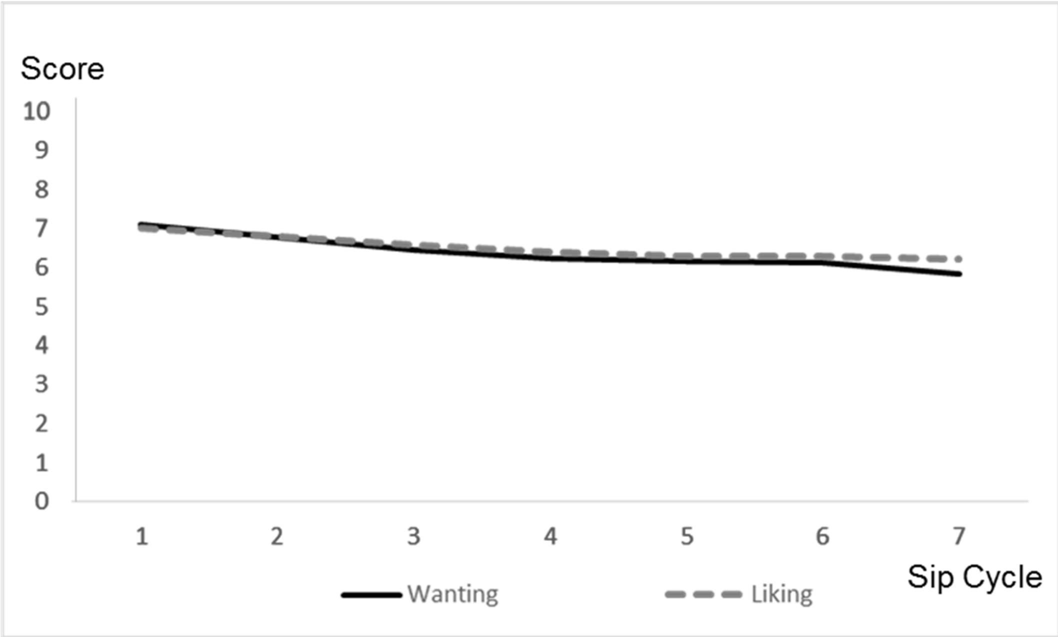

Figure 8a. TDS difference curves of the samples B4 and B3 – Naïve panel.

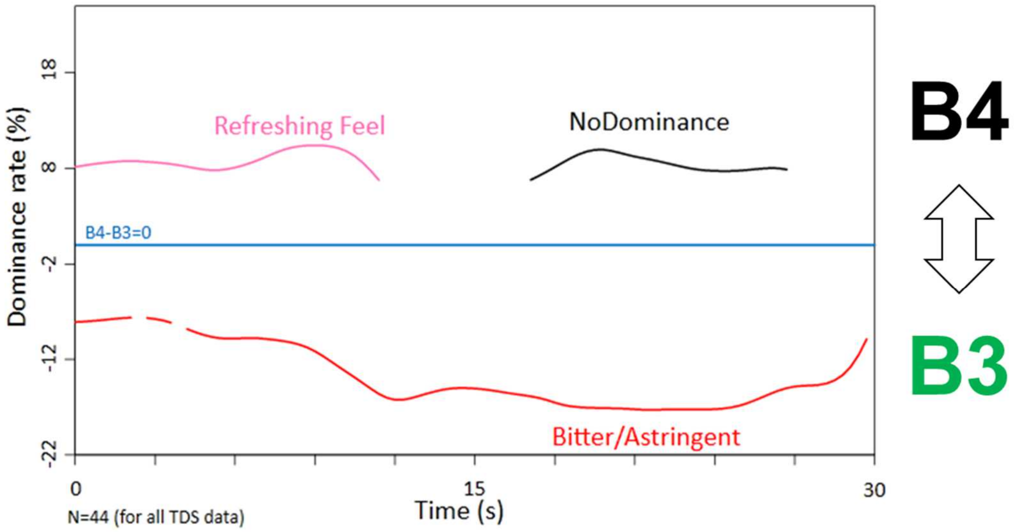

Figure 8b. TDS difference curves of the samples B4 and B1 – Naïve panel.

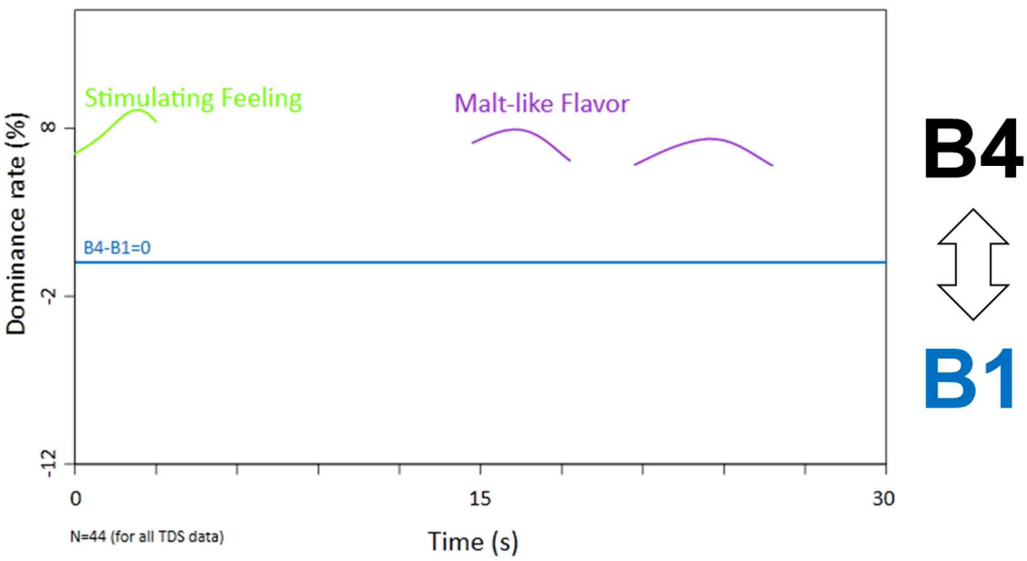

Figure 9. TDS difference curves of the samples B4 and B1 by wanting segmentation – Naïve panel.

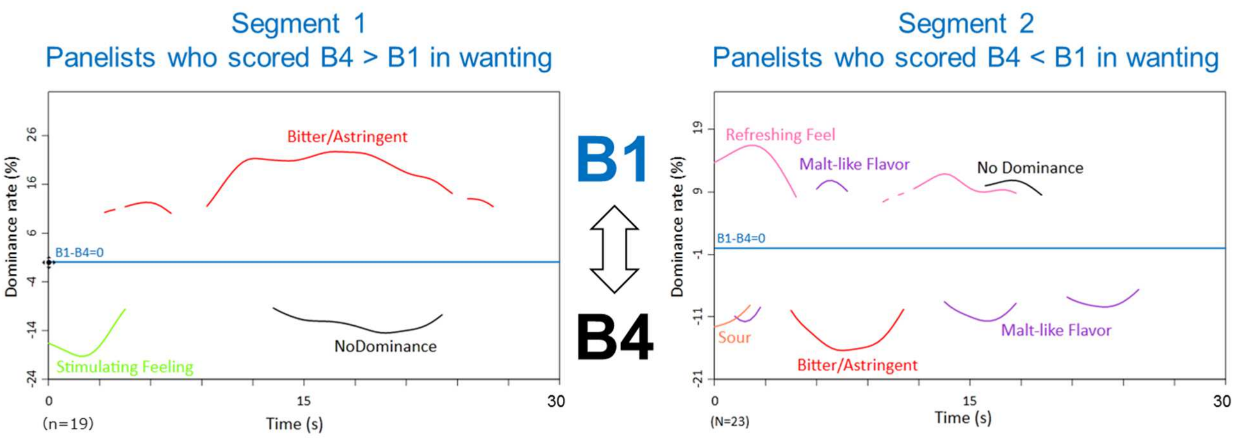

Figure 10. Development of Duration: *Bitter/Astringent* - Naïve Panel

Dominance duration means with their 95% confidence level.

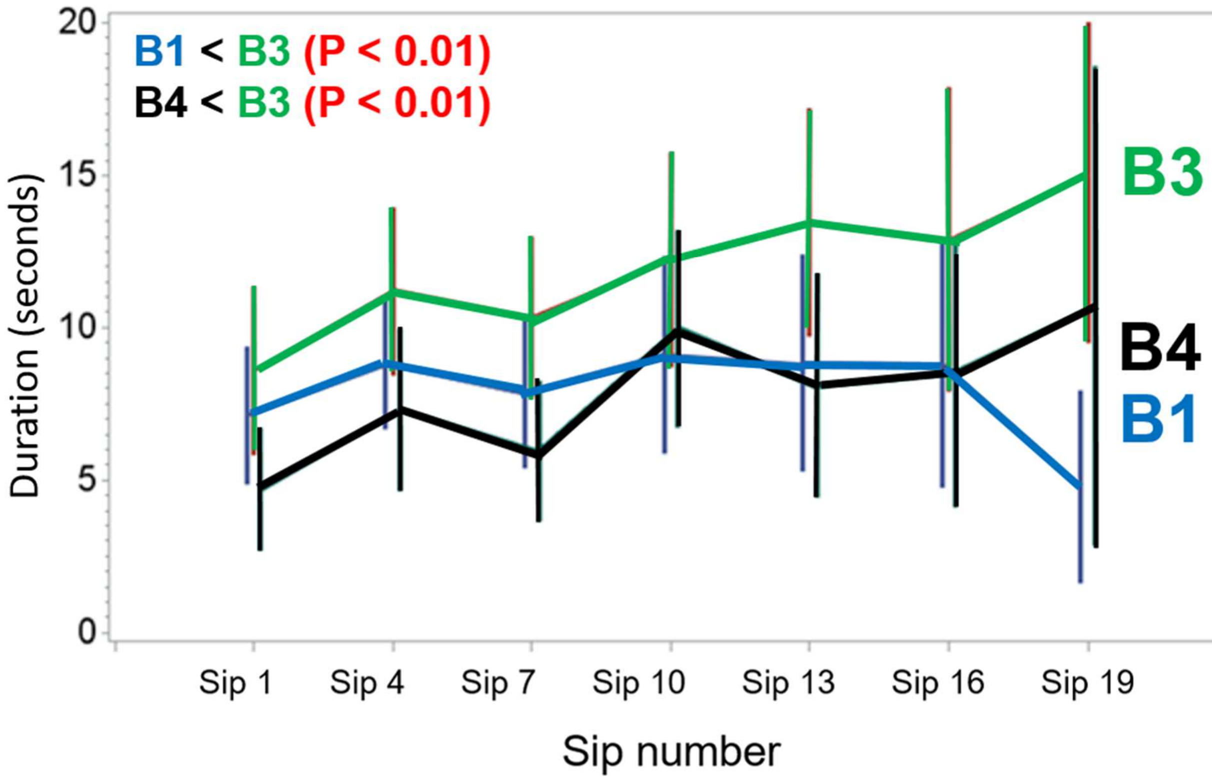

Figure 11a. Volume of beer actually consumed by the expert panel (Mean ± Standard Error x 2).

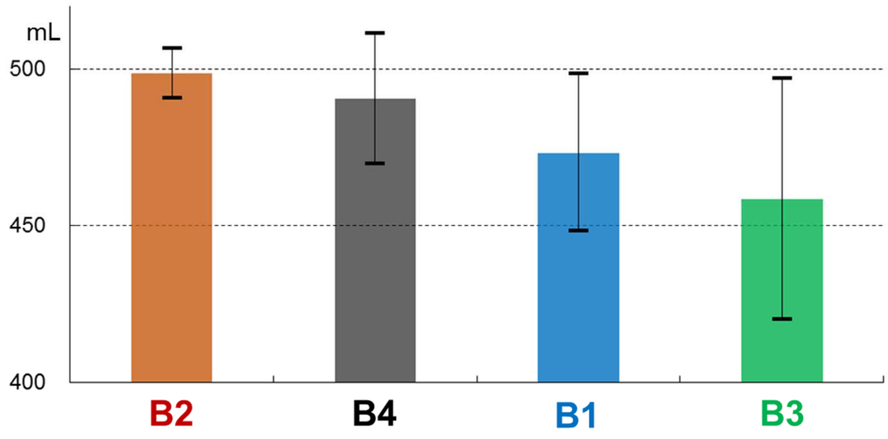

Figure 11b. Volume of beer actually consumed by the naïve panel (Mean ± Standard Error x 2).

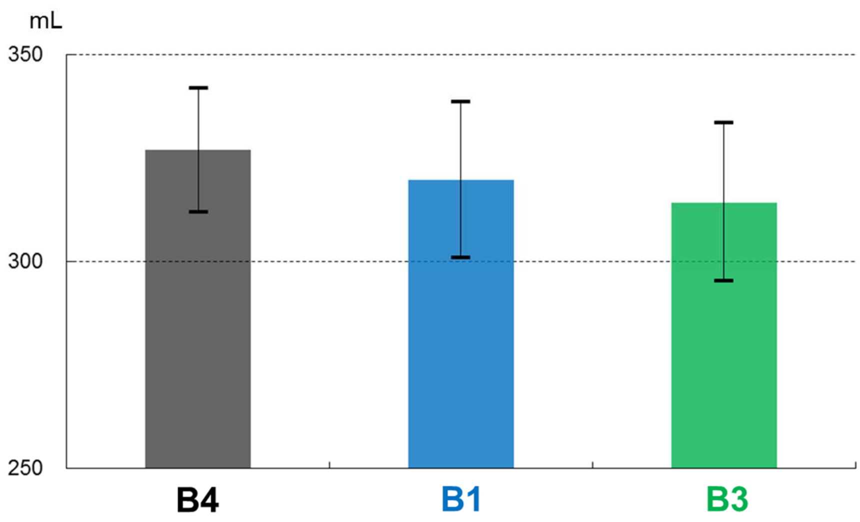

Table 1

ABV, IBU, Linalool and sensory descriptions of beer samples used in this study.

Sample	Alcohol content (% ABV)	Bitterness (IBU)	Linalool (ppb)	Sensory descriptions
B1	5	22.5	21.3	*Hoppy, Malty, Bitter*
B2	5	19.5	1.5	*Less standout flavors, Light finish*
B3	5	29	2.4	*Malty, Bitter, Full body*
B4	5	19.5	1.2	*Less standout flavors, Light finish*

(Note: ABV = alcohol by volume. IBU = international bitterness unit. ppb = parts per billion.)

(Note: Information about alcohol content was obtained from the product package. Bitterness and linalool were analyzed in the R&D center at Asahi Breweries, Ltd. Sensory descriptors were obtained from the results of quantitative descriptions and sensory comments, both of which were acquired from the Asahi expert beer sensory panel.)

Table 2a

ANOVA of dynamic wanting, liking, ease of drinking and residual sensation: Repeated Mix Model (1/2) - Expert Panel -
Using Restricted Maximum Likelihood (REML) algorithm with first-order heterogeneous auto-regressive covariance structure among sips

	p-values of tests of fixed effects			Product LS-Means			
Descriptor	Sip	Product	Product*Sip	B1	B2	B3	B4
Wanting	0.0000	0.0037	0.0310	3.65a	5.95b	3.78a	5.59b
Liking	0.0023	0.0155	0.2672	4.46ab	6.29b	4.26a	5.89b
Ease of Drinking	0.0172	0.0003	0.6351	3.88a	6.62b	3.83a	5.98b
Residual Sensation	0.2290	0.4738	0.5048	4.18a	3.98a	4.79a	3.65a

Means with two different letters are significantly different ($p < 0.05$).

Table 2b

ANOVA of dynamic wanting, liking, ease of drinking and residual sensation: Repeated Mix Model (2/2) - Expert Panel -
Using Restricted Maximum Likelihood (REML) algorithm with first-order heterogeneous auto-regressive covariance structure among sips

	Sip Cycle LS-Means							p-values of Sip Effect	p-values of Sip Effect by Product			
Descriptor	1	2	3	4	5	6	7		B1	B2	B3	B4
Wanting	6.18	5.67	5.35	5.15	4.71	4.23	2.99	0.0000	0.0000	0.2367	0.0000	0.1576
Liking	6.02	5.78	5.36	5.26	5.09	4.69	4.92	0.0023	0.0008	0.9161	0.0084	0.2998
Ease of Drinking	5.85	5.67	5.46	5.17	5.08	4.71	3.84	0.0172	0.2475	0.5702	0.0869	0.1434
Residual Sensation	4.18	3.74	3.80	3.60	3.65	4.14	4.93	0.2290	0.5085	0.8558	0.0447	0.4299

Table 3a

ANOVA of dynamic wanting, liking and ease of drinking: Repeated Mix Model (1/2) - Naïve Panel -
Using Restricted Maximum Likelihood (REML) algorithm with first-order heterogeneous auto-regressive covariance structure among sips

	p-values of tests of fixed effects			Product LS-Means			p-values for comparing products		
Descriptor	Sip	Product	Product*Sip	B1	B3	B4	B1-B3	B1-B4	B3-B4
Wanting	<0.0001	<0.0001	0.1767	6.52b	5.69a	6.95b	<0.0001	0.1007	<0.0001
Liking	0.0040	<0.0001	0.5063	6.74b	5.70a	7.11b	<0.0001	0.1346	<0.0001
Ease of Drinking	0.0041	<0.0001	0.6762	6.80b	5.66a	7.23b	<0.0001	0.1037	<0.0001

Means with two different letters are significantly different ($p < 0.05$).

Table 3b

ANOVA of dynamic wanting, liking and ease of drinking: Repeated Mix Model (2/2) - Naïve Panel -
Using Restricted Maximum Likelihood (REML) algorithm with first-order heterogeneous auto-regressive covariance structure among sips

	Sip Cycle LS-Means							p-values of Sip Effect	p-values of Sip Effect by Product		
Descriptor	1	2	3	4	5	6	7		B1	B3	B4
Wanting	7.09	6.79	6.46	6.23	6.15	6.14	5.85	<0.0001	0.0101	0.0001	0.1893
Liking	7.03	6.81	6.58	6.40	6.28	6.28	6.24	0.0040	0.1616	0.0084	0.4305
Ease of Drinking	7.03	6.82	6.67	6.51	6.40	6.42	6.11	0.0041	0.4283	0.0172	0.1261