

Stability in consumer responses to familiar and new chocolates during a period of exposure.

Tormod Næs*+, Anne Segtnan*, Britt Signe Granli *, Elena Menichelli*, and Margrethe Hersleth*-

*** Nofima, Oslovegen 1, 1430 Ås, Norway**

***+ Dept. of Food Science. Univ. Copenhagen, Rolighedsvej. Faculty of Science, Denmark**

- Department of Chemistry, Biotechnology and Food Science, The Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

Abstract

The aim of this paper is to study the effects of exposure to different types of chocolate on consumer responses. Three different chocolates were considered, one established market leader and two new chocolates from a product development project. A group of consumers participated two times in a central location test and were exposed to different chocolates over a 4 weeks period between the two occasions. Both average effects and individual differences in consumers' responses are discussed. The main conclusion of this study is that the evolution of hedonic liking during product exposure may be dependent on initial liking of the products. It is important to further improve acceptance testing strategies for food.

Key words: Exposure, familiarity, liking, individual differences, product development

1. Introduction

One of the main purposes in experimental consumer studies of food is to obtain information about consumers' hedonic liking or purchase intent for a number of products of interest (Lawless & Heyman, 2010). Typically a group of representative consumers (usually between 100 and 200) are invited to participate either in a Central Location Test (CLT) or in a Home-Use Test (HUT), and each of them is asked to give a personal assessment of the presented products. In some cases, the consumers are asked to rate the products using a scale anchored with "dislike extremely" and "like extremely" (Lawless & Heyman, 2010; Peryam & Pilgrim, 1957) while in other cases they are asked to rank the products according to preference or to choose the most preferred products from a number of choice sets (Gustafsson, Herrmann, & Hubert, 2003; Næs, Brockhoff, & Tomic, 2010). Recently, a number of new approaches have been proposed in the literature for the purpose of obtaining more realistic consumer data. Examples are various types of experimental auctions (Lange, Martin, Chabanet, Combris, & Issanchou, 2002) and studies based on so-called take-away strategies (Næs, Monteleone, Segtnan, & Hersleth, 2013; Weiss, O'Mahony, & Wichchukit, 2010; Wichchukit & O'Mahony, 2011). These tests are different from the standard acceptance testing as they monitor what people actually do, not the scoring of acceptance or preferences for products. Information from all these different types of studies is important for a product developer in order to decide which decisions to take before a product launch.

Another aspect of interest for a product developer is the dynamic evolution in consumers' responses during exposure of a product, i.e. to what extent do single hedonic measurements have a predictive value for future liking and choice? To answer this question several studies have been undertaken where the consumers are asked several times during exposure about their assessment of the products (Koster, Couronne, Leon, Levy, & Marcelino, 2003; Köster, Rummel, Kornelson, & Benz, 2001; Mustonen, Hissa, Huotilainen, Miettinen, & Tuorila, 2007; Kinneer & Kock, 2011; Stolzenbach, Bredie, Christensen & Byrne, 2013; Sulmont-Rossé, Chabanet, Issanchou & Köster, 2008). Measurements of hedonic flexibility (Mustonen et al., 2007) and indices of performance (Næs et al., 2013) have also recently been developed for the purpose of measuring these aspects at an individual level. Measurement of acceptance for a product over time is particularly relevant when a new product is to be launched in a market where corresponding and competitive products exist. There are very few published papers which address such an issue; a possible reason being that often this is addressed within an industrial, confidential product development context. Some recent examples can be found in Stolzenbach et al, (2013) and Kinneer and de Kock (2011).

The aim of this paper is to compare consumers' responses over time for an established chocolate in the market with two new chocolates in the same product category. Consumers' responses for the chocolates were collected during a period of four weeks, starting and ending with a Central Location Test (CLT). In the two CLTs the consumers were asked about hedonic liking and they were also asked to indicate their choice of chocolate for take-away

after that last CLT was finished. Between the two CLTs selected consumer groups were exposed to the three different products and during this exposure period the consumers were asked to rate hedonic liking for this product three times. In this paper we will compare consumers' acceptance ratings before and after exposure as well as their take-away choice when they are given the possibility to choose products for bringing home. General population trends as well as individual differences in the different product assessments will be studied.

2. Materials and methods

2.1 Chocolate samples

This study is linked to a large industry based product development project (conducted in 2011). The purpose was to compare consumer's responses to three chocolates (without any filling) in a given product category. One chocolate was the established and leading brand (E) and the two other ones were new chocolates (N1 and N2). (There was one more chocolate in the original test, but due to production problems, this was taken out from all external reporting). The two new products were produced in a pilot plant by the collaborating industrial partner in this project, while the established brand from a competitor had to be purchased in a supermarket.

The established brand samples were re-melted before the test in order to make the size and visual appearance of the pieces as similar as possible to the other two samples.

2.2 Consumers

Two hundred consumers were recruited from voluntary associations in the Oslo region. Each consumer was paid for the participation, but honorariums were given directly to the association. Thus, the consumer had no personal benefits from participating in the test except some chocolate to bring home. The consumers were between 20 and 60 years old and they reported a chocolate consumption of twice a week or more. In total 193 of the 200 consumers consumed the established brand most often as compared to the most important competitor in the same chocolate category.

2.3. Design of the study

The design of the study is shown in Figure 1. Only the responses indicated in bold are discussed in the paper. These consumer responses (except familiarity) were based on informed rating and informed choice, which would be responses most comparable to consumers' perception in real situations when eating and/or choosing a product. Familiarity of the products was also included in CLT1 as such a parameter could contribute to understand the evolution of consumers' hedonic liking of the three products during exposure. The familiarity was scored on a scale from 1-9 for which 1 was not known at all and 9 was very well known. The same scale between 1 and 9 was used for informed liking. The rest of the data were used for industrial purposes only

The test consisted of two CLTs named CLT1 and CLT2, with a 4 weeks exposure period between. In CLT1 the respondents were presented samples of the three chocolates and asked

to rate blind liking, familiarity (blind), perceived complexity (blind), expected liking (informed), informed liking and finally to rank the three samples (informed) to bring home after CLT2 (take-away test). The consumers were told that after CLT2 they would be given one large bar of one of the chocolates, but since we were uncertain about the selection of products available, the consumers were asked to rank the three products according to choice (for take-away after the test). In the informed test, information about brand was given to the consumers (competitors brand for the established and own brand for the two new products).. For the two new products the information intended for printing on the final chocolate packaging was given. This information included communication of sensory properties, for N1: “a chocolate with a distinct flavor of cocoa and sweetness” and for N2: “a powerful and rich chocolate”.

The consumers were allocated to four different groups of N=50 for the exposure period. Consumers in three groups (G1, G2 and G3) were given one of the three chocolates (E, N1 and N2) to bring home, while the consumers in the fourth group (G4) did not get any chocolate to bring home. In other words, there were three consumer groups exposed to different products and one reference group. The chocolates that were given for exposure at home were branded in the same way as in the CLT's

For the exposure period each consumer was given 3 bars of 200 gram chocolate. The consumers were encouraged to eat the chocolate regularly during the 4 weeks (preferably every second day). They could share the chocolate with their family, but they should regularly eat some chocolate themselves. Three times during exposure (one time every week) the consumers received an e-mail from Nofima, in which they were asked to indicate informed hedonic liking, the amount of chocolate eaten and to report intake of other chocolates since last reporting.

After 4 weeks (30 days) the consumers came back to CLT2. During this test they were again presented samples of the three chocolates for rating blind liking, informed liking and to rank the three samples (informed) for final choice to bring home. As already stated, only data from familiarity, informed hedonic rating and choice are discussed in this paper.

The samples are randomized in the program EyeQuestion by Logic8 BV 2001-2012 in a balanced design.

In order to make the ranking and rating comparable in the plot, the ranks were transformed to (4 minus the rank) so that a rank 3 means the most preferred.

2.4. Statistical analyses

2.4.1. Average effects

For analyzing informed liking for CLT1, a standard ANOVA with effects for product (fixed) and consumer (random) is used. When exposure group is involved, this will be included as an additional fixed factor. In this case, the interactions between group and product will also be incorporated and studied. For this model, consumer (random) is nested within exposure group. In order to be able to compare the two occasions we will use differences in rating between

CLT2 and CLT1. The same models as above will be used. We will present both p-values, and main effects plots.

The home test results will be analyzed by using simple graphs and ANOVA for each of the three groups separately. The model will contain effects of time period and consumer.

The take-away ranks and changes of the ranks will be analyzed by graphs and plotting of the changes between the two occasions (differences). These will be compared with the corresponding results for rating data. A Friedman test will also be used for comparing ranking at the two occasions using differences between the two tests. The effects will be product and consumer.

2.4.2 Individual differences

For investigating individual differences, tables comparing ratings before and after exposure will be made. A 2-way table of % increased/decreased liking with rows for the products and columns for the different exposure groups will be presented. Histograms of the same differences for the three products will be presented.

The take-away test and the rating test will be compared to investigate general consistency in consumers' assessments of the three products. This will be done by computing standard deviations over the four assessments for each consumer and product. These standard deviations will be plotted in histograms for each product.

3. Results.

3.1 Effects of exposure

3.1.1 Initial liking values for rating and take-away (CLT1)

The ANOVA results, both for the simple model with only consumer and products effects and for the model with group effect incorporated are presented in Table 1. As can be seen, the product effect is strongly significant. The group effect and the group*product effects are, however, non-significant, which is to be expected since this response is given before exposure and the splitting into groups is done randomly. The effect sizes are presented in Figure 2. As can be seen, the established product (E) received the highest informed liking. The three products are all significantly different in liking (Tukey's test). Product N2 received the highest liking of the two new chocolate samples.

The plot of the average take-away ranks is shown in Figure 3 and illustrate a close correspondence (on average) between consumers' ratings and their choice (ranking) for the take-away test (see also Næs et al(2013) for similar results).

3.1.2 Comparing average ratings in all tests

The general average tendencies for all the three exposure groups are presented in Figure 4. As can be seen, the hedonic liking increases in the home test period compared to the CLT1 for all three products. During the home test period the ratings are very stable. The standard errors

for the home test ratings varied between 0.14 and 0.22. After the home test, the rating stays constant for the two new products, while it goes down by more than one unit for G1 (the E group).

The average results and their confidence intervals for CLT2 are given in Figure 5. As can be seen, the average pattern is different for the four groups. The G1 group, who got exposed to the established product and G4, who had no exposure are quite similar with slightly higher average values for G4 (reference group). For G2 (exposed to sample N1) there were no difference between products, while for the G3 (exposed to sample N2), the liking is highest for the N2 sample. Compared to Figure 2, the exposure had a positive effect on the liking of the new products, but not for the established one. In order to compare the two CLTs statistically, the same type of ANOVA as above was done for the differences CLT2-CLT1 (see Table 2). As can be seen, all effects are significant at 5% level. The plot of the corresponding differences is given in Figure 6 together with the confidence intervals. The significance tests for the products within each group showed that there are significant differences between the products for G2 (exposed to N1), in which product E is significantly different from N1 and N2. For the G1 (exposed to E) and G4 (reference group) groups no significant product effects were found at 5% level, while for G3 (exposed to N2) the p-value was close to significant at 10% level ($p=0.11$). The product N1 in G2 increases by about 0.8 units in liking and product N2 in G3 by about 0.3 units as compared to CLT1.

The corresponding results for the take-away choices are shown in Figure 7. As can be seen, the results are very comparable to the results from hedonic liking (Figure 6). The Friedman non-parametric tests gave the following p-values for the 4 groups (0, 97, 0.07, 0.06, 0.07, all adjusted for ties). The only slight difference from the ratings is that differences are slightly larger for the reference group, with an increase in ranking value for N1.

In general, the informed liking increased for the two new products (N1 and N2) and decreased for the well-known product (E) during this experiment. In G2 and G3, the increase in liking was highest for the product the consumer was exposed to. In other words, exposure improved liking for the two new products, in particular this effect was evident for N1 which had the lowest initial liking. For the other two groups (G1 and G4), the situation was stable, no significant differences between the products during the experiment.

3.1.4. Individual exposure effects for rating and take-away

The number of consumers, represented as percentages, that increase/decreased their liking (or have the same liking) for the products they were exposed to are represented in Table 3. The table is organized according to exposure group and product as above, i.e. products as rows and columns as groups. Histograms for the differences in liking between CLT1 and CLT2 for product E in G1 (the E exposure group), product N1 in G2 (the N1 exposure group) and product N2 in G3 (the N2 exposure group) are presented in Figure 8. It is clear that for G1 (exposed to sample E) the liking values go more in the negative direction (less liked in the second test) than in the positive direction. For G2 (exposed to N1), the situation is the opposite. For the G3 (Exposed to N2) there is a slight positive tendency. These tendencies correspond to the average results in Figure 6.

3.2. Stability/flexibility of the consumers over the different assessments

The average differences between rating and ranking are discussed above. Here we go more in detail on the individual data. For the comparison in both CLT1 and CLT2 the rating values were transformed into ranks. In the case of ties, the two rating values are given the same rank. The following results will be given in terms of percentages of equal ranks for the three categories “best liked”, “next to best liked” and “liked least”. This will be done for each product separately and given in the order in the previous sentence.

The actual %-values of equal rank in CLT1 for product E were 82, 41, 52, for product N1 the corresponding values were 40, 49 and 72 while for product N2 they were 37%, 62% and 51%. In CLT2, for product E the %-values were 84, 58, 66, for product N1 they were 63, 50 and 77 and for product N2 they were 45, 76 and 66. As can be seen, for the best liked product the values are clearly the highest for product E (82% and 84%). Among the other two products, the “best liked” had the lowest (or close to lowest) consistency between the rating and ranking. In other words, for the market leader (product E) the values are very consistent, while for the two new products (N1 and N2) they are quite inconsistent for the best liked product.

4. Discussion and conclusions

The results in this paper have shown that for the well-established product (E), the exposure period had no positive effect on hedonic liking, except for a temporary increase during the home test. The positive tendency during home exposure could be due to a more realistic tasting situation (context), which has been demonstrated to influence liking scores (Hersleth, Mevik, Næs & Guinard, 2003). However, the reduction in rating for E between the last reporting of liking in the exposure period (HTW3 in Figure 4) and CLT2 is less easy to explain. Actually as many as 193 out of the 200 consumers were regular consumers of this brand. It was therefore not expected that the positive informed liking of this product should change during the exposure period.

For the two new products, the ratings during the home test period were quite stable, but significant increases in ratings were found in CLT2 compared to CLT1 for N1. Effect of exposure has been studied in the literature mostly for blind tasting conditions. One exception is a study by Stolzenbach et al. (2013). In this study effects of repeated exposure were studied for informed apple-juices over 3 weeks including 10 exposures. No effect of exposure was found for the apple-juices in this study. It was concluded that the juices were resistant to change in sensory perception and that concept associations over repeated consumption were considered “exclusive” even though they were considered familiar. This implies that in this case the increased familiarity of the product during the exposure period did not have any effect on ratings.

Effects of exposure on hedonic rating in blind tasting conditions shows different results according to the arousal potential of the products (see e.g. references in Sulmont-Rosse et al, 2008) Sulmont-Rosse et al. (2008) studied the role of stimulus arousal potential in the effect of repeated exposure on the liking for fruit drinks. Results showed that exposure led to an

increase in liking for drinks with a high arousal potential, while no evolution of liking was observed for drinks with a moderate arousal potential. Moreover, Zandstra, Weegels, Van Spronsen, and Klerk (2004) concluded that a positive or negative effect of repeated exposure might depend on the initial level of novelty of the stimulus. It is therefore of interest to discuss the results in this chocolate study in light of the results from the familiarity rating described in Figure 1. These results show that familiarity was quite similar for all three products with averages in the range between 5.1 and 5.6 (in CLT1) on a scale between 1 and 9 with N1 having the lowest and N2 the highest familiarity. In other words, the least familiar product in CLT1 got the highest increase in liking after exposure, which to a certain extent corresponds to the results in Sulmont-Rosse et al (2008) and Zandstra et al, (2004). (Kinnear and de Kock (2011) found effects of product exposure in a study on sports drinks.. Products that were initially preferred became less preferred after 14 days. The consumers in this study consumed sport drinks relatively frequently and were probably familiar with the products.

Regarding the responses for informed choice, the increased ratings found for the two novel products from CLT1 to CLT2 were confirmed by the measurement of choice for take-away (ranking). Similar results were found in Næs et al. (2013) for cured ham. At an individual level, there were, however, large differences among the consumers in their assessments.

Based on the results in this study it may be concluded that evolution of reported informed liking during an exposure period may depend on the initial liking and/or the initial novelty of the product, although the degree of novelty in this case was only weakly visible in the average familiarity value for one of the new products. Accordingly, it can be difficult to predict future hedonic liking of food products based on one single experimental testing session. This confirms earlier experiences and conclusions stated by among others Kinnear & de Kock, 2011 and Kremer et al, 2013. It is therefore highly important to consider the initial familiarity of the products during planning, organizing and interpretation of results from consumer testing. Moreover it is also important to further improve acceptance testing strategies for food.

Acknowledgements. We would like to thank the research Council of Norway and the Foundation of Research Levy on Agricultural Products in Norway for funding of this project (Food Choice project and Consumer Check project).

Table 1. ANOVA tables for the three products in CLT1 (for both with and without group effect)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
product	2	185,173	185,173	92,587	34,11	0,000
consumer	199	1024,240	1024,240	5,147	1,90	0,000
Error	398	1080,160	1080,160	2,714		
Total	599	2289,573				

Source	DF	Seq SS	Adj SS	Adj MS	F	P
group	3	7,215	7,215	2,405	0,46	0,708
product	2	185,173	186,406	93,203	34,09	0,000
group*product	6	8,518	8,518	1,420	0,52	0,794
consumer (group)	196	1017,025	1017,025	5,189	1,90	0,000
Error	392	1071,642	1071,642	2,734		
Total	599	2289,573				

Table 2.Anova of differences between CLT2 and CLT1 results

Source	DF	Seq SS	Adj SS	Adj MS	F	P
group	3	39,214	39,214	13,071	2,77	0,043
product	2	42,070	45,429	22,715	6,21	0,002
group*product	6	54,133	54,133	9,022	2,47	0,024
consumer (group)	196	926,384	926,384	4,726	1,29	0,018
Error	392	1434,463	1434,463	3,659		
Total	599	2496,265				

Table 3. Percentages of lower, equal and higher in test 2. The four groups are presented as columns and the 3 products as row. For E in the G1 group, the number of consumers that go down in liking is quite high (44%) and for the N1 in G2 it is low (14%). For N2 in G3 it is in between.

	G1	G2	G3	G4
E	44/30/26	41/35/24	41/44/15	35/30/35
N1	32/34/34	14/29/57	39/22/39	38/24/38
N2	46/24/30	20/29/51	28/39/33	33/29/38

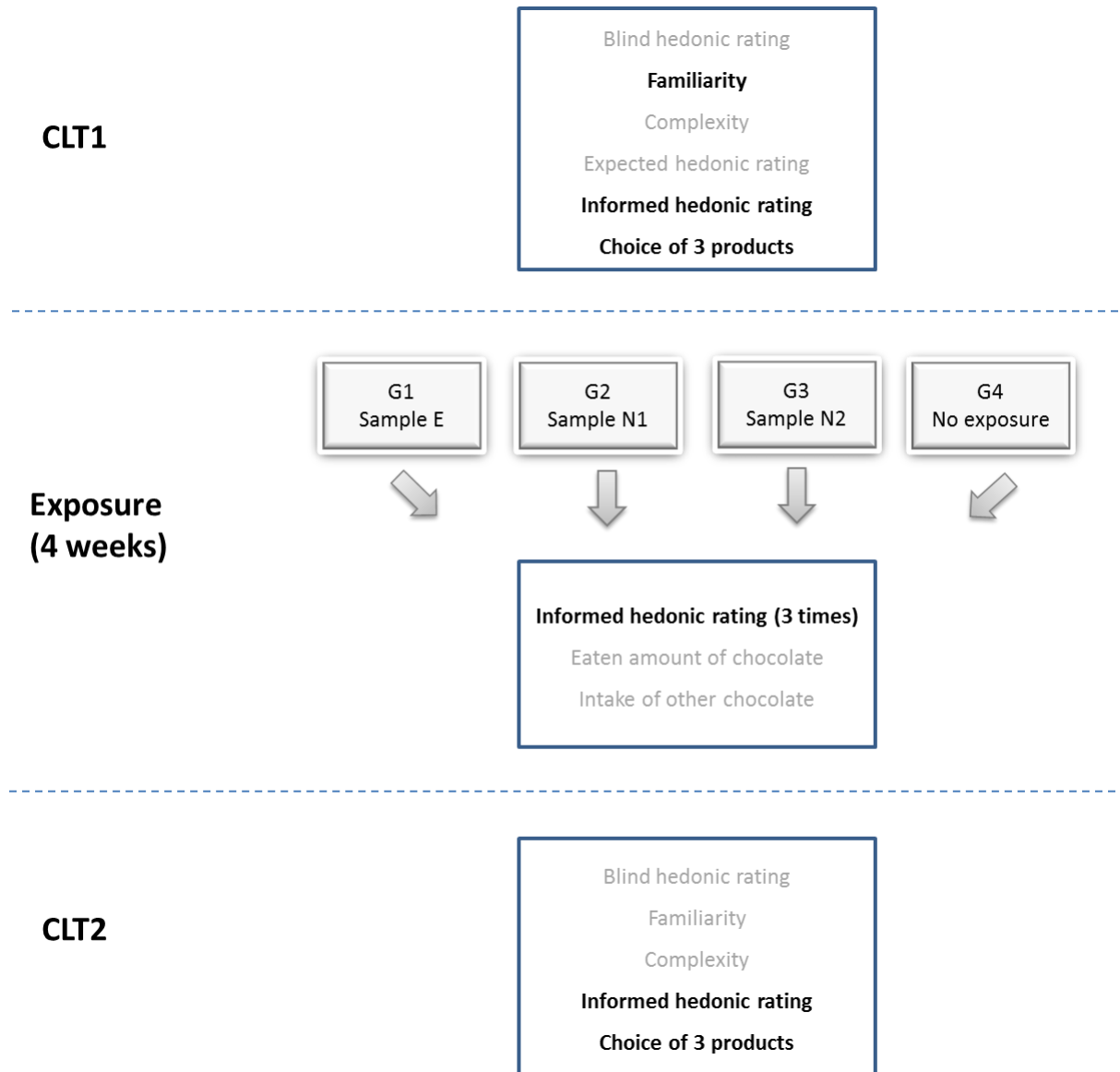


Figure 1. A graphical illustration of the experiment

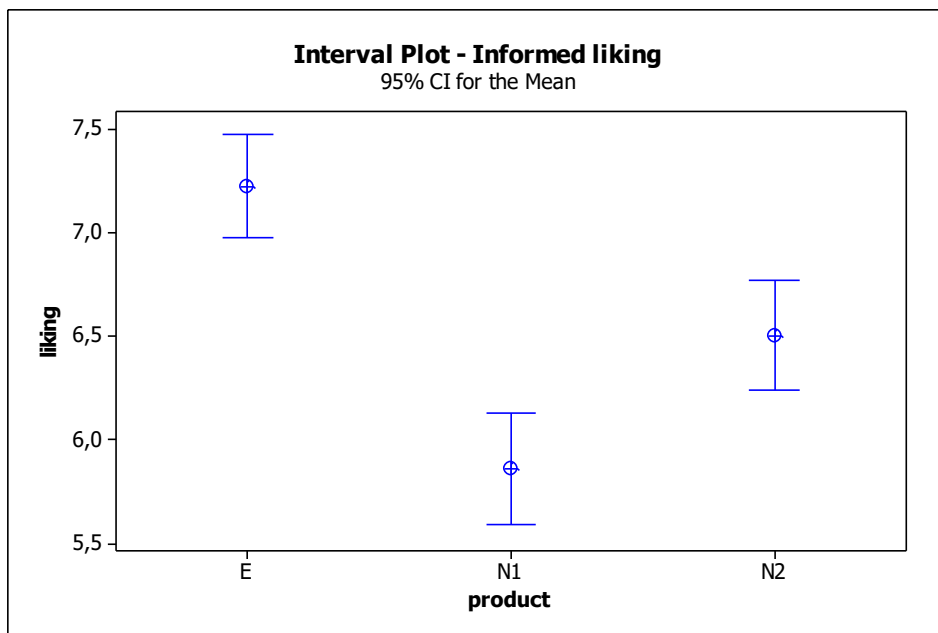


Figure 2. Average informed liking in CLT1.

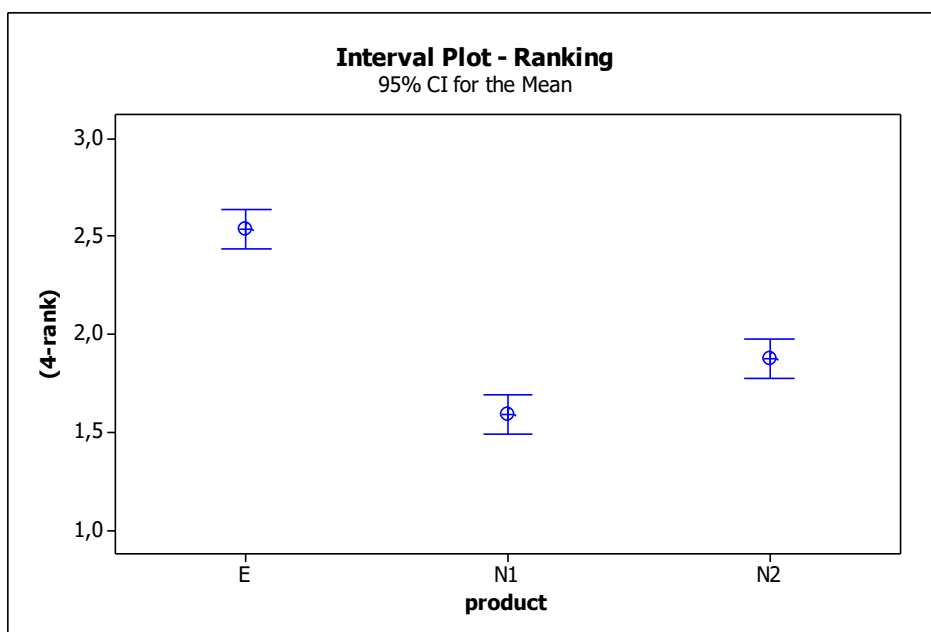


Figure 3. Average informed take-away ranks in CLT1.

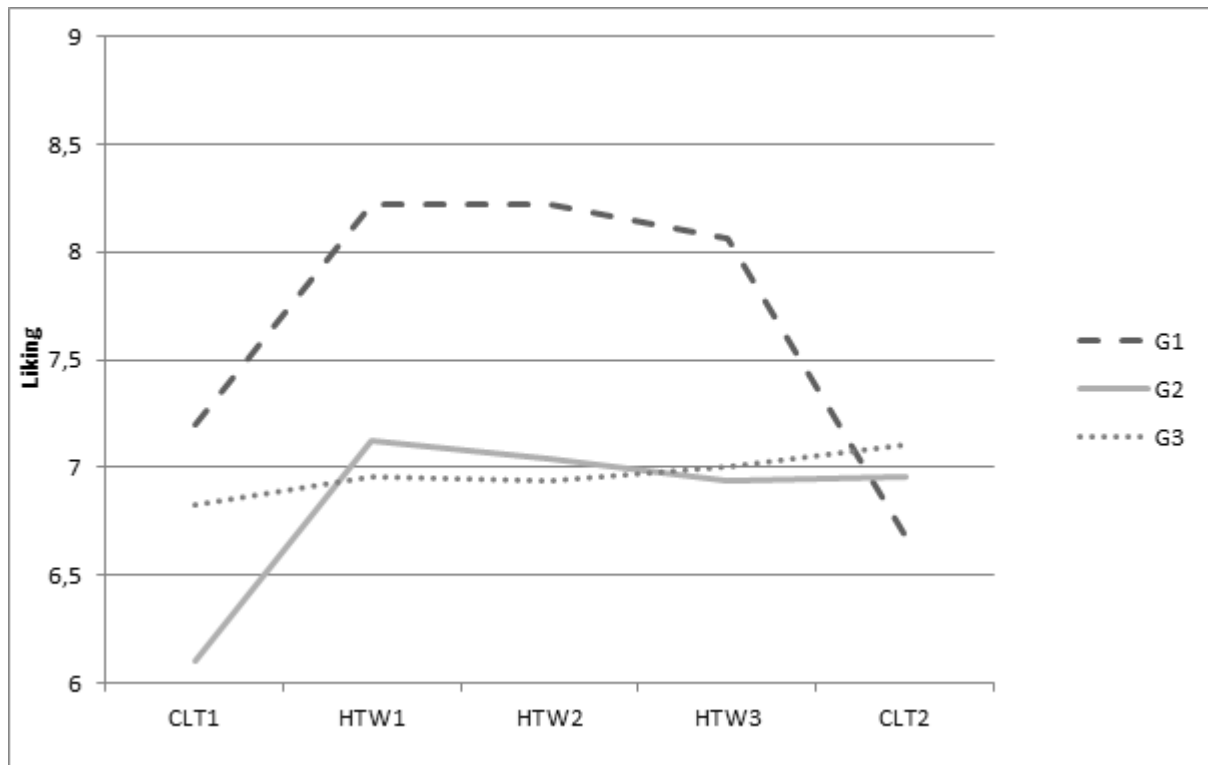


Figure 4. Averages for the CLT's and the home tests. The standard errors for the home test averages are between 0.14 and 0.22.

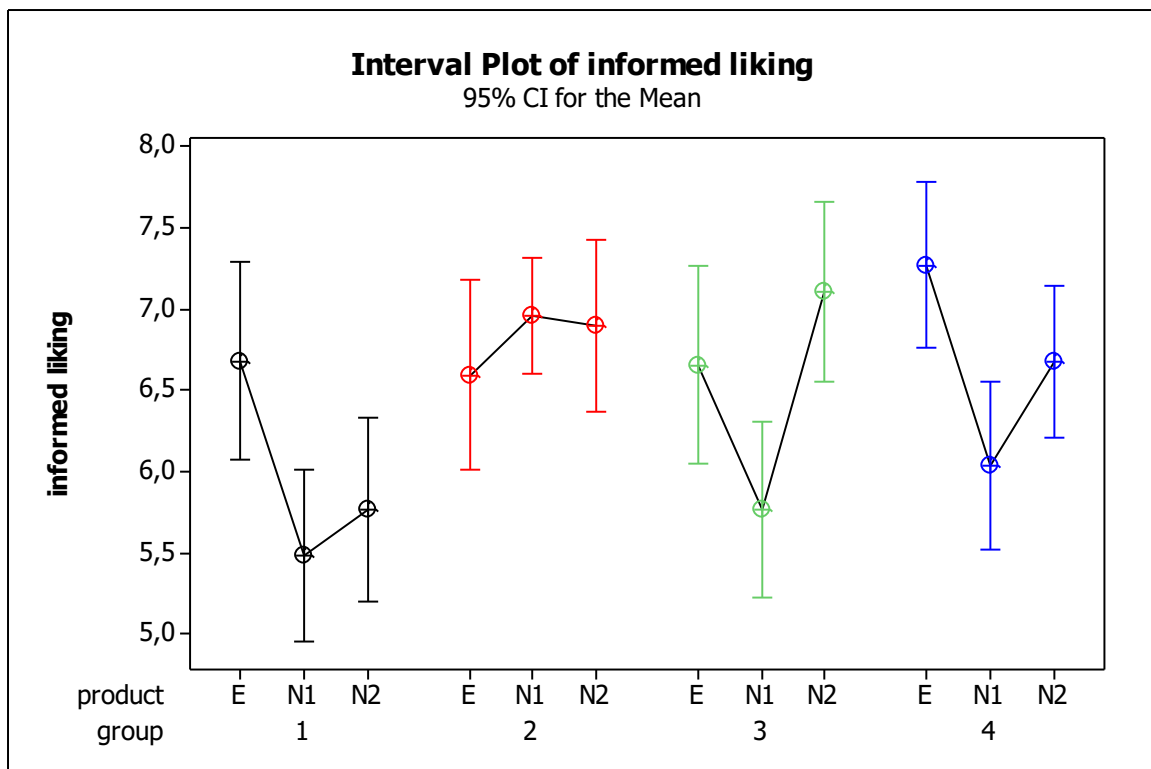


Figure 5. Average results for CLT2 in the four groups.

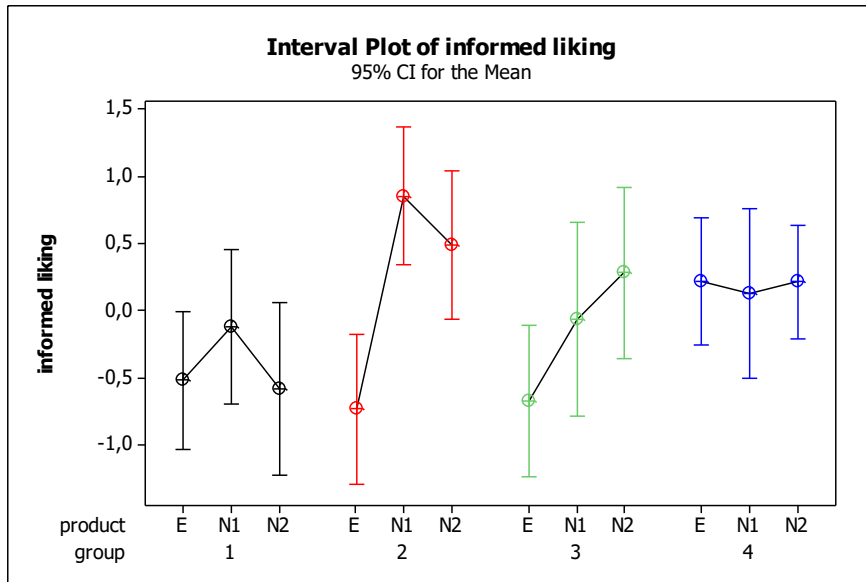


Figure 6. Differences in informed liking for the four exposure groups. Differences taken as CLT2-CLT1 values for each product.

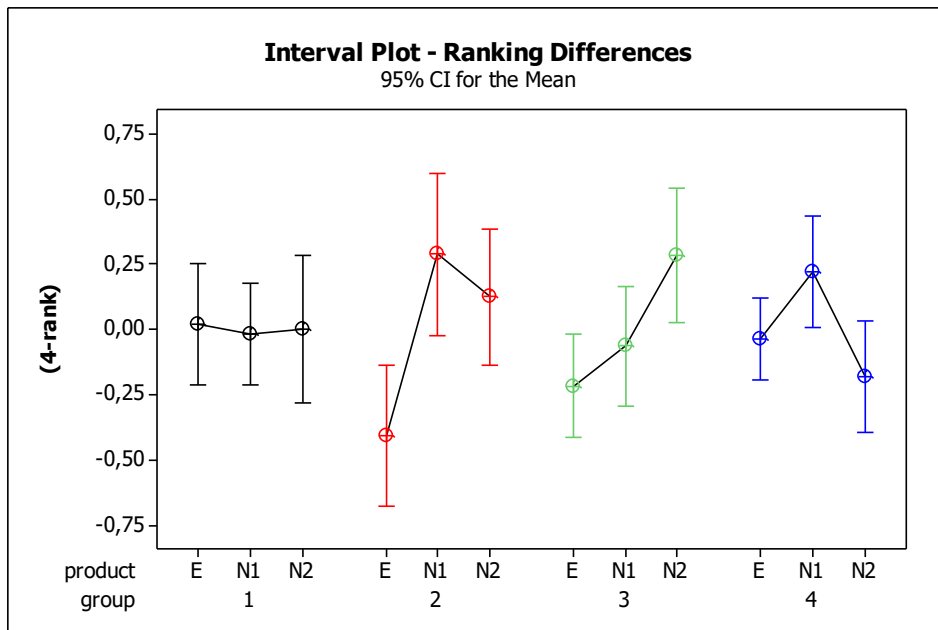


Figure 7. Average differences of take-away ranks in the four exposure groups. Differences taken as CLT2-CLT1 values for each product.

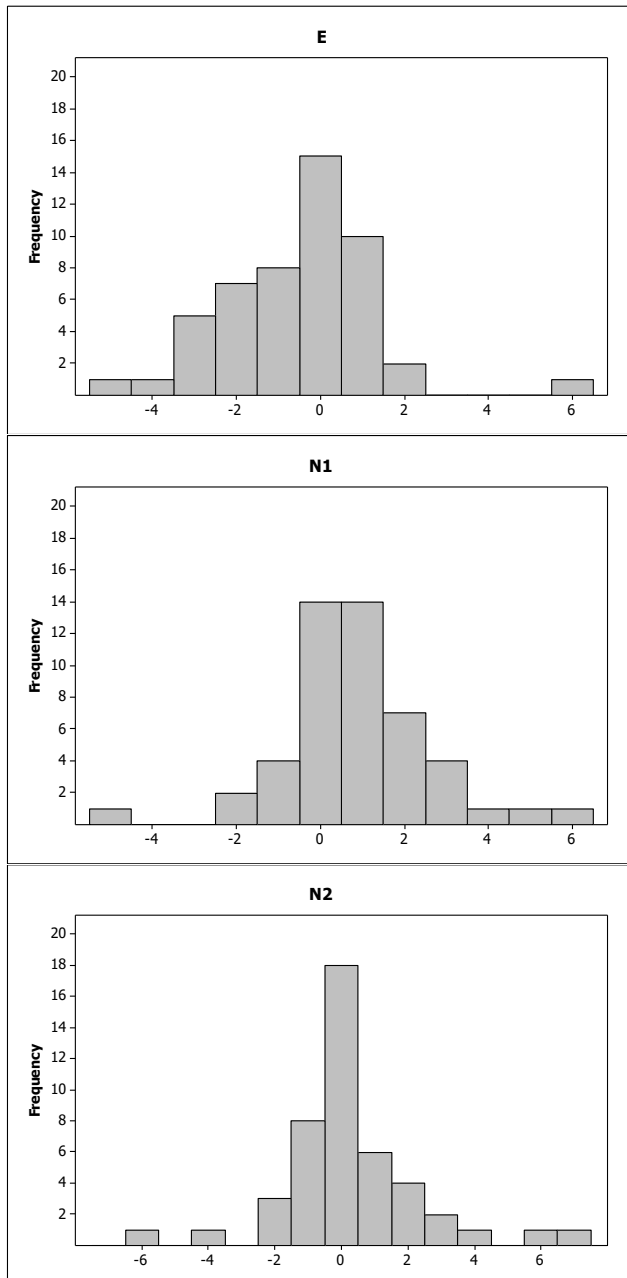


Figure 8. Histograms of the differences in liking before and after exposure for the three samples (in its own exposure group). Positive values correspond to higher values in the CLT2.

References

- Gustafsson, A., Herrmann, A., & Hubert, F. (2003). *Conjoint measurement - Methods and Applications*. Berlin: Springer Verlag.
- Hersleth, M., Mevik, B.-H., Næs, T., Guinard, J.-X. (2003). Effect of contextual factors on liking for wine - use of robust design methodology. *Food Quality and Preference*, Vol 14, 7, pp 615-622.
- Kinnear, M., & de Kock, H. L. (2011). Would repeated consumption of sports drinks with different acidulants lead to hedonic adjustment? *Food Quality and Preference*, 22(4), 340-345.
- Koster, E. P., Couronne, T., Leon, F., Levy, C., & Marcelino, A. S. (2003). Repeatability in hedonic sensory measurement: a conceptual exploration. *Food Quality and Preference*, 14(2), 165-176.
- Kremer, S., Shimojo, R., Holthuysen, N., Koster, E. P., & Mojet, J. (2013). Consumer acceptance of salt-reduced "soy sauce" bread over repeated in home consumption. *Food Quality and Preference*, 28(2), 484-491.
- Köster, E. P., Rummel, C., Kornelson, C., & Benz, K. H. (2001). Stability and change in food liking: food preference in the two Germany's after the reunification. In M. Rothe, *Flavour 2000: perception, release, evaluation, formation, acceptance, nutrition and health*. Germany: Rothe, Bergholz-Rehbrücke.
- Lange, C., Martin, C., Chabanet, C., Combris, P., & Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for Champagne: comparison with hedonic scores. *Food Quality and Preference*, 13(7-8), 597-608.
- Lawless, H. T., & Heyman, H. (2010). *Sensory evaluation of food. Principles and practices*. New York: Springer.
- Mustonen, S., Hissa, I., Huotilainen, A., Miettinen, S.-M., & Tuorila, H. (2007). Hedonic responses as predictors of food choice: Flexibility and self-prediction. *Appetite*, 49(1), 159-168.
- Næs, T., Brockhoff, P., & Tomic, O. (2010). *Statistics for sensory and consumer science*. Chichester, UK: Wiley.
- Næs, T., Monteleone, E., Segtnan, A., & Hersleth, M. (2013). A comparison of two new take-away strategies and their relation to rating and ranking of extrinsic properties of dry cured ham. *Food Quality and Preference*, 27(1), 63-71.
- Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology*, 11(9), A9-A14.
- Stolzenbach, S., Bredie, W. L. P., Christensen, R. H. B., & Byrne, D. V. (2013). Impact of product information and repeated exposure on consumer liking, sensory perception and concept associations of local apple juice. *Food Research International*, 52(1), 91-98.
- Sulmont-Rosse, C., Chabanet, C., Issanchou, S., & Koster, E. P. (2008). Impact of the arousal potential of uncommon drinks on the repeated exposure effect. *Food Quality and Preference*, 19(4), 412-420.
- Weiss, B. H., O'Mahony, M., & Wichchukit, S. (2010). Various paired preference tests: Experimenter effect on "take away" choice. *Journal of Sensory Studies*, 25(5), 778-790.
- Wichchukit, S., & O'Mahony, M. (2011). 'Liking', 'Buying', 'Choosing' and 'Take Away' preference tests for varying degrees of hedonic disparity. *Food Quality and Preference*, 22(1), 60-65.
- Zandstra, E. H., Weegels, M. F., Van Spronsen, A. A., & Klerk, M. (2004). Scoring or boring? Predicting boredom through repeated in-home consumption. *Food Quality and Preference*, 15(6), 549-557.