Book of abstracts

Sensometrics2020

Keynote speakers

Quantitative Intuition: Combining prior knowledge and big data

Harald Martens[1,2], Andreas Wulvik[1] and Silje Skeide Fuglerud[3]

[1]Idletechs AS, Trondheim Norway (harald.martens@idletechs.com). [2]Dept Engineering Cybernetics, NTNU Trondheim Norway. [3]St.Olavs University Hospital, Trondheim Norway

Clear, logical thinking is just the tip of the iceberg. Intuition, first impression, gut feeling: This is the brain on autopilot, a result of unconscious mental parallel processing of many types of input at the same time. The mind does some sort of multivariate analysis, without being fully aware of it. Our beliefs, values, preferences are multivariate summaries of past experiences, cumulative knowledge, peer pressure, culture.

The body does multivariate analysis, too, and brilliantly so. Otherwise we could not ride a bike. Our heart beats are beyond our control, like other cyclic processes in our biology. Our breath is a link between body and mind – half physical, half mental. Modern measurements show our brain hard at work. Drinking a double espresso, perhaps just thinking of that espresso, and all changes.

Over time, the classical data in sensometrics – 3rd person sensory perception reports – e.g. descriptive sensory panel data and consumer report - are increasingly supplemented by technical observation data, from cameras, microphones etc as well as physiological continuous streams of high-dimensional data from body measurements by EEG, ECG, skin resistance, FMRI, blood- or saliva analyzers etc. How to make sense of this new torrent of multichannel data?

For each given setting, there is usually a *limited* number of independent causalities behind the observational data. Some of these causalities are already well known. We can specify their effects in advance in terms of e.g. mathematical models, and estimate their state variables by fitting these models to the stream of empirical data.

But the data will usually be affected by other, unexpected causalities too, whether we like it or not. These unexpected and unmodelled variation patterns may create so-called alias errors – they interfere with the quantification of the known causalities - unless corrected for. Moreover, these new patterns may be very informative in their own right.

With the right software tools for multivariate hybrid modelling, you can discover these unexpected, but clearly varying causalities empirically, as systematic patterns in the multichannel observational data. The known and the unknown causalities may then be interpreted simultaneously, in light of your tacit background knowledge, to eliminate the alias problems and gain new insight. Here we demonstrate how some of the standard methods for multivariate data modelling in sensometrics have been modified and extended, to yield tools for quantitative intuition. In other words: Explainable AI in practice.

See also: https://www.camo.com/iatl_haraldmartens/

Segmentation with complex data: Arriving at an insightful representation

Marieke E. Timmerman
Psychometrics and Statistics, Heymans Institute for Psychological Research, University of Groningen, the Netherlands; m.e.timmerman@rug.nl

Consumer segmentation aims at dividing consumers into groups of individuals that are similar in specific ways, as expressed in their empirical data. In this talk I will present mixture modelling as a useful cluster analysis approach to achieve an insightful segmentation. As mixture modelling is highly versatile, it can deal with complex data, giving the opportunity to define similarity in various ways. Herewith, it is useful to distinguish different types of variables (e.g., continuous, ordinal), and different data structures. I will discuss various prototypical data structures, including single-set and multi-set multivariate data, longitudinal data and multilevel data. The associated potentially useful models include mixture factor models, growth mixture models, and Markov models. The models will be presented from a birds eye view. Then I will turn to empirical practice, and discuss how to arrive at a proper model for an empirical segmentation problem at hand. As I will discuss, key steps are to identify the data structure and the aspects to cluster on. I will present different exemplary empirical cases from psychology, to show the approach and its usefulness for achieving insight into subgroups of individuals.

Controversy regarding relevance and rigor of Sensometrics for industrial applications

Michael Meyners
Procter & Gamble Service GmbH, 65824 Schwalbach am Taunus, Germany

What makes a method from sensometrics (not) to be adopted and used in industry? When does "quick and dirty" become a bit too dirty? How many different methods do we need to address the same question? Or aren't they addressing the same question?
Academic research and industry needs do not always closely align. Many (sensory and statistical) methods are developed and deployed, but they are rarely compared exhaustively and objectively with alternative existing methods. Why would I adopt any new method when I have something that currently (seemingly) addresses the same task in a similar way? What benefit does it bring, and is it important enough for me to bother? How can I make findings actionable to inform product design?
Simplified and faster methods and their related analyses have become popular over the last decade, with at best anecdotal evidence that results are comparable to classical, often more laborious efforts. Simultaneously, research is conducted that seems to suggest that lower efforts (i.e. fewer panelists, fewer replications, less training, …) may still give similar results. Is similar good enough? Was the standard itself good enough to begin with, or might it already have degraded over a few previous iterations of cost optimizations (think about the famous salami slicing)? What gold standards do we adhere to in order to avoid degradation of data quality over time?
This presentation aims at triggering discussions about what we are doing well today and what opportunties exist. It will become a plea to strengthen both the relevance and statistical rigor across the discipline. If we continue on our current path, Sensometrics might deteriorate its own credibility.

Session 1- Individual differences and consumer segmentation – part 1

O1.1 Combining hedonic information and CATA description for consumer segmentation: new methodological proposals and comparison

Evelyne Vigneau[1], Véronique Cariou[1], Davide Giacalone[2], Ingunn Berget[3], Fabien Llobell[4,1]
[1]Oniris, StatSC, Nantes, France. [2]University of Southern Denmark, Odense, Denmark. [3]NOFIMA, As, Norway. [4]Addinsoft, XLSTAT, Paris, France

Check-all-that-apply (CATA) has become a popular method for obtaining a consumer-based sensory characterization. Moreover, in some situations, consumers are also asked to evaluate the set of products according to a liking scale as well as CATA questions, with the aim to identify the key sensory attributes associated with the most liked, or disliked, products.

While consumer segmentation has been widely studied in the context of hedonic data, the integration of CATA information when identifying consumer segments is still a recent subject. Most of the studies combining liking scores and sensory CATA description, consist, first, in the identification of consumer segments based on the preference profiles. The second step consists in the analysis of the CATA responses, usually based on the cross tabulation matrix containing total frequency of mention for each attribute.

Our purpose is to investigate several alternatives for *simultaneously* identifying clusters of preference profiles while taking into account the CATA attributes. These alternatives are derived from strategies already proposed by the different co-authors, with various modifications either in terms of data preparation or algorithm development. These alternatives are to:

a) Apply Fuzzy Clusterwise Regression (FCR)[1] for identifying groups of consumers with similar relationship between CATA and liking data;
b) Use the Clustering around Latent Variables (CLV) approach for the panel segmentation based on the matrix of the liking scores while taking account of aggregated CATA description as external data[2];
c) Combine liking and CATA information at the individual level and performing a CLUSTCATA-like[3] procedure with a new similarity index;
d) Like c), combine the two kind of information at the individual level leading to a three-way array thereafter partitioned with the CLV3W[4] approach.

The different strategies will be illustrated using real datasets such as the rye-bread dataset[5]. The results will be compared on the basis of indices encompassing interpretability, cluster stability and configuration aspects.

[1] Wedel, M., & Steenkamp, J.-B. E. M. (1991). *Journal of Marketing Research*, 28(4), 385–396.
[2] Vigneau, E., Endrizzi, I. and Qannari, E. M. (2011). *Food Quality and Preference*, 22: 705-713.
[3] Llobell, F., Cariou, V., Vigneau, E., Labenne, A. and Qannari, E.M. (2019). *Food Quality and Preference*, 72, 31–39.
[4] Cariou, V., and Wilderjans, T. F. (2018). *Food Quality and Preference*, 67, 18-26.
[5] Giacalone, D. (2018). Product Performance Optimization. In Ares, G., & Varela, P. Methods in Consumer Research, Volume 1 (pp. 159-185), Elsevier.

O1.2 Do consumers focus on the same terms in a CATA task?

Thierry Worch
Friesland Campina, Wageningen, Netherlands. Qi Statistics, Theale, United Kingdom

In consumer research, the CATA task has gain popularity due to its simplicity and quickness to perform, as assessors are asked to tick from a list the words that would characterize the samples they are testing.

From a data point of view, CATA tasks translate into binary variables, the variable taking a '1' if the word has been ticked by the assessor for the given sample, and a 0 otherwise. Univariate and multivariate analyses can be performed on such data: For example, the data can be aggregated across assessors to build a sample x word frequency table, which is then used to map the associations between samples and words using multivariate techniques such as Correspondence Analysis.

But since the contingency table is obtained by aggregating the assessors' responses, individual differences are ignored, and some valuable information may be lost in the process: Are all the assessors focusing on the same terms?

In this presentation, a new methodology for clustering consumers based on their association terms – samples is proposed. This methodology measures the similarity between pairs of assessors using Dice's criterion, hence helping better understanding the panel of consumers by grouping those that are ticking the same set of words, and by separating them from those who have other association of interest. Additionally, this methodology can also be adapted to the words themselves, hence providing sets of words that tend to be ticked simultaneously, and separated from terms that are not.

To illustrate the methodology, different CATA tests involving sensory terms, emotion terms, concepts/benefits etc. will be considered. Thanks to these different studies, examples showing how this methodology can be used to show 1) which criterion are important for consumers, 2) which terms tend to be ticked simultaneously, and 3) which type of data would be more suitable for such clustering procedure will be discussed.

O1.3 Determination of the number of clusters of subjects in Projective Mapping, Free Sorting and CATA experiments

Fabien Llobell[1,2], Evelyne Vigneau[1], El Mostafa Qannari[1], Véronique Cariou[1]
[1]StatSC, ONIRIS, INRA, Nantes, France. [2]Addinsoft, XLSTAT, Paris, France

Projective mapping, Free Sorting and Check-All-That-Apply (CATA) are rapid methodologies for sensory characterization that involve untrained or semi-trained subjects. Thus, when analysing the data obtained by means of such approaches, it is of paramount of interest to assess whether the panel corresponds to a homogeneous group, or contrariwise, if subgroups of subjects could be identified.

The data from projective mapping, free sorting and CATA tasks, can be represented as multiblock datasets, with each block corresponding to a subject. Clustering techniques of multiblock data such as CLUSTATIS [1] and CLUSCATA [2] aim at simultaneously clustering subjects and computing, within each cluster, an average group configuration that could be used to depict the relationships among the products. Notwithstanding, as with any clustering method, the choice of the appropriate number

of clusters is a of paramount importance. The aim of this work is to provide statistical tests and criteria to answer the two important following questions: (1) is it appropriate to segment the panel of subjects?, (2) if so, how many clusters are there?

To answer these questions, different approaches have been investigated on the basis of simulated as well as real datasets. The two approaches that turned out to be the most efficient were implemented in the R package ClustBlock [3]. The first approach, which aims at answering the first question, starts by computing a similarity matrix between the subjects (*e.g.* RV coefficients) and, thereafter, investigates the significance of the second eigenvalue of this matrix. In order to answer the second question, we propose an adaptation of the Hartigan index [4] to the multiblock case.

References

[1] Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2020). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. Food Quality and Preference, 79, 103520.
[2] Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. Food Quality and Preference, 72, 31-39.
[3] Llobell, F., Vigneau, E., Cariou, V., Qannari, E. M. (2019). ClustBlock: Clustering of Datasets. R package version 2.1.0. https://CRAN.R-project.org/package=ClustBlock
[4] Hartigan, J. A. (1975). Clustering algorithms. John Wiley & Sons.

O1.4 Influence of ticking style on the validity of CATA data with 6-9-year-old children

Martina Galler[1,2], Paula Varela[3], Valérie L. Almli[1], Tormod Næs[1]
[1]Nofima, Ås, Norway. [2]NMBU, Ås, Norway. [3]Nofima, ÅS, Norway

CATA (Check-all-that-apply) is now an established method to generate quantitative product descriptions in a simpler way than traditional methods such as Quantitative Descriptive Analysis (QDA). Many studies have shown that the application with consumers can generate valid results by e.g. comparison with the QDA of a trained profiling panel. Examples are described in Varela and Ares (2012). Due to the simplicity, CATA is promising for consumer studies with special populations such as children. Several recent publications such as Laureati et al. (2017) describe the application of the method for product testing with children.

In this paper, two case studies on bread and smoothies (samples based on DOE) were used to assess how children perform CATA, and how individual differences in the responses to the test can affect the validity of the test results.

The discrimination capability was evaluated as first indicator for the validity of the data. Additionally, a second validation indicator, the comparison to the QDA of a trained profiling panel, was considered. To understand how children use the CATA list, their ticking style as suggested by Næs, Varela, and Berget (2018) was evaluated considering different indicators: average number of ticks per sample (low vs. super tickers), variance of the number of ticks per sample (steady vs. unsteady tickers), the number of CATA attributes used (few vs. many different CATA terms) and the average variance per attribute (indicator for low vs. high sample discrimination). The assessors were grouped according to their ticking style and their results were compared regarding discrimination between samples and the similarity to the sensory description by the QDA of a trained profiling panel .

The discussion of the results will cover methodological aspects on how the children perceived the samples, how the sensory space reflected the liking space, and the influence of the test protocol between the two case studies and how these may influence the results.

Laureati, M., Cattaneo, C., Lavelli, V., Bergamaschi, V., Riso, P., & Pagliarini, E. (2017). Application of the check-all-that-apply method (CATA) to get insights on children's drivers of liking of fiber-enriched apple purees. *Journal of Sensory Studies, 32*(2). doi:10.1111/joss.12253
Næs, T., Varela, P., & Berget, I. (2018). *Individual Differences in Sensory and Consumer Science, Experimentation, Analysis and Interpretation*: Woodhead Publishing.
Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*(2), 893-908. doi:10.1016/j.foodres.2012.06.037

Session 2: Textmining and network analysis

O2.1 Crowdsourcing consumer research: Understanding the importance of restaurant atmosphere via text mining and sentiment analysis of restaurant reviews in a large Yelp dataset.

Qian Janice Wang

Aarhus University, Aarhus, Denmark

It is well-known that multisensory store atmospherics, such as background music or smell, can alter consumer behaviour in the retail world. However, there is limited empirical research investigating atmospherics in a restaurant setting. This is partly explained by the difficulty of conducting onsite restaurant surveys and the expenses involved in manipulating restaurant atmosphere.

Recently, the availability of large bodies of data from social media combined with natural language processing (NLP) techniques makes it possible to conduct consumer research in a novel and highly relevant way. More specifically, crowd-sourced restaurant reviews present the opportunity to understand consumer trends across a broad spectrum of populations and restaurant categories.

The present study focuses on the relevance of restaurant atmospherics, specifically auditory atmosphere, in the dining experience. Sound is an often overlooked aspect of the dining experience, although there is increasing evidence that what we hear can influence food liking and flavor evaluation, not to mention purchase behavior. In the present study, I investigated the instances in which the auditory atmosphere was mentioned in a restaurant review, and analysed the utility and emotional content of such reviews.

To this end, a large dataset was analysed from the 2019 Yelp Dataset Challenge, providing 224,858 reviews in 69,081 restaurants over ten metropolitan areas in North America. Text mining identified 12,426 reviews which mentioned background music or sound level. An analysis of restaurant categories revealed that bars and modern cuisine restaurants had the highest proportion of sound-related reviews. Moreover, compared to reviews in general, sound-related reviews were peer-rated as most useful in hotels, bakeries, and ethnic restaurants.

Sentiment analysis was conducted using AFINN lexicon, available in the tidytext R package, in order to classify a given review in terms of positive/negative affect. Results demonstrated that, relative to all reviews, sound was reviewed in the most positive context in wine bars and tapas restaurants, where as fast food, dance clubs, and hotels had the most negative sound-related reviews.

The present study supports the use of natural language processing methods as a tool in conducting consumer research, in order to identify and capture consumers trends and preferences from the wealth of crowd-sourced data available.

O2.2 Sorting Backbone Analysis: A network-based method of extracting key actionable information from free-sorting task results

Jacob Lahne Virginia Tech, Blacksburg, VA, USA

The free-sorting task is increasingly popular as a rapid sensory method to give a global picture of the similarities among samples. Sorting does not require training analysts, allows for the easy, simultaneous presentation of up to 20 samples, and provides stable results with 25-30 subjects. However, wide use of free-sorting is hindered by the current analyses for free sorting—for example DISTATIS and Correspondence Analysis—which require statistical expertise to conduct and interpret. In this paper a novel, alternative analysis is proposed, called “Sorting Backbone Analysis” (SBA), which is based on tools from network analysis. The similarity data produced from free sorting can represent a weighted network, and so a set of network-analysis tools can be used to identify groups of products which are significantly similar, and to visualize these results clearly and powerfully. SBA is simple and costs nothing to implement, provides interpretations that agree with current methods, and produces clear, powerful visualizations called “graphs,” which are much simpler to interpret than the standard biplots familiar to sensory scientists. This paper describes the mathematical and statistical background for SBA and applies SBA to six, previously published sorting datasets, with comparisons to DISTATIS. In each case SBA produces visual results that highlight all of the same features as the standard approach while being easier to interpret, and in many cases produces new insights. Therefore, SBA specifically and network analysis in general are suggested as new approaches for use in the analysis of sensory similarity data as produced through free sorting and related methods.

O2.3 How to Use a Graph Database to Extract Insights from Diverse Historical Sensory and Consumer Data: A Step-By-Step Use-Case

Alex M. Pierce-Feldmeyer[1], John M. Ennis[2]

[1]Mane, Inc., Lebanon, USA. [2]Aigora, Richmond, USA

This talk presents a step-by-step use-case of how to implement a graph database to extract novel insights from diverse historical sensory and consumer data. Among the types of data we include in a single database are: Acceptance, demographics, descriptive analysis, difference testing including both Tetrad and Triangle tests, emotion CATA, ingredient and flavor information, Just-About-Right, preference, and temporal data. We demonstrate how, using Neo4j - a freely available open-source graph database platform - sensory scientists can perform efficient and intuitive artificial intelligence

(AI) powered queries on the totality of their historical data. We then explain how these queries would be difficult to perform using more traditional relational database technology.

Keywords: Artificial intelligence, graph databases, big data, historical data

O2.4 Why Graph Databases Are the Answer to the Question, "What Should We Do with Our Historical Data?"

John Ennis[1], Alex Pierce-Feldmeyer[2]

[1]Aigora, Richmond, USA. [2]Mane, Inc., Lebanon, USA

Sensory and consumer science teams typically possess vast quantities of historical data, usually in a wide variety of formats and collected using a wide variety of methodologies. In addition, these same teams typically have access to similarly vast quantities of instrumental and analytical data as well as to data concerning ingredients and formulations. In some fortunate cases, these teams even have access to consumer behavioral data such as sales and complaint records. But how should these teams organize this information so that the resulting database can be used to support new research efforts? In this talk, we answer this question by explaining how graph databases are ideally suited to capture the many relationships between the disparate information contained in these many datasets. We review the concept of a data model and introduce a particular category of data model called "response-centered" that allows for facile and rapid access to precisely the information required to investigate new research questions using historical data. We explain why graph databases overcome the limitations of traditional relational databases that might historically be used for the same purposes, and demonstrate how attendees can use the freely available open-source graph database platform Neo4j to begin warehousing and querying their historical data.

Keywords: Artificial intelligence, graph databases, big data, historical data

Session 3: Individual differences and consumer segmentation – part 2

O3.1 Mixed Assessor Model for Scheffe type paired comparison data

Hironori Satomura

Tobacco Science Research Center, Japan Tobacco Inc., Yokohama, Japan

Scheffe type paired comparison is defined as a task in which each assessor rate degree of sensory differences between the presented pair of the products with using predefined attributes. Typically, data from Scheffe method is analyzed by analysis of variance with using a special design matrix. However, as pointed out by various authors for the analysis of monadic sensory profiling data, linear models including analysis of variance are not capable of properly handling scale range differences between individual assessors. This study first explains why modeling scale range differences is (more) important for Scheffe type paired comparison (than for the standard profiling test); then, proposes a data analysis method that explicitly models scale range differences between individual assessors. This can be viewed as a generalization of the mixed assessor model. A procedure to estimate

parameters of the model is derived. To explain how the proposed procedure works and to illustrate its usefulness, some numerical results are presented. Finally, possibility of further extensions of the method will be discussed.

O3.2 Sensory profile optimization through preference distribution prediction for target demographics and consumer cohorts

Yizhou Ma, Jason Cohen, Ryan Ahn

Analytical Flavor Systems, New York, USA

Measurement of consumer preference in sensory science often involves hedonic scores, and product developers rely on hedonic scores to determine consumer acceptance of products. Hedonic scores are commonly reported as mean-liking and variance, which implicitly assumes that consumer preference of a product is normally distributed.

However, the distribution of a product preference may not be normally distributed. Understanding the specific preference distribution of a product can provide additional decision metrics for a product and further help product developers to optimize its sensory profile.

By sampling consumer preference for on-market products and labeling the data with demographic and tasting-experience scores, a quantitative predictive model of consumer preference can be developed on a per-demographic and per-cohort basis. This model can predict the distribution of product preference for specific target consumer demographics or specific cohorts. The predicted preference distribution can guide sensory profile optimization towards various consumer cohorts within the same demographic, leading to more preferred products targeted to each cohort.

A demographic baseline preference model was developed by collecting consumer preference of 5 to 20 on-market food and beverage products from 800 unique individuals in the United States (US). The baseline preference model was trained using Bayesian additive regression trees and validated by minimizing the cross entropy of the prediction via multinomial log-loss. In this study, a common cocktail “Manhattan” made from the standard formulation was reviewed by US consumers with various demographic backgrounds and tasting experiences. The preference distribution of standard “Manhattan” was predicted from a baseline preference model using Bayesian inference of the hedonic score distribution. The predicted preference distribution was further segmented by various consumer cohorts to identify preference differences of the standard “Manhattan”.

Flavor profile optimization of the standard “Manhattan” was conducted towards different segments within the overall consumer population. Formulations were identified by optimizing the standard “Manhattan” based on targeted demographic cohorts. The predicted preference distributions of the optimized formulations were validated by measuring hedonic scores from the targeted demographics. An alignment of the predicted and observed preference distribution showed that more than one preferred “Manhattan” formulation exists among US consumers based on demographic backgrounds or consumer cohorts within a demographic. Product sensory profile can be better optimized for targeted demographics and clusters of consumers within each demographic by predicting their preference distributions.

O3.3 Global Data Set Segmentation: The Impact of Our Pre-Clustering

Michael Gasho[1], Juan Martinez[2], Patti Wojnicz[2]

[1]PepsiCo Sensory and Consumer Product Insights, Plano, TX, USA. [2]P&K Research, Chicago, IL, USA

Background: The use of cluster analysis on liking data to determine preference segments is common in sensory science. The mechanics of forming data clusters are typically algorithm based with specific rules governing the final clustering solution. These rules are blind to the nature of the incoming data and rely on the data analyst to appropriately transform (e.g. center, rescale, etc.) the data beforehand. This transformation is becoming increasingly important as barriers to survey respondents and collect global data sets decrease; especially sets which employ the 9-point hedonic scale where significant by-country scale usage behavior differences exist.

Methods in Progress: A 6000 respondent, 12 country online survey was conducted where a standard screening of respondents were asked their opinion of 8 food concepts on the 9-point hedonic scale. Histograms & descriptive statistics were developed on the aggregated liking responses on a *per country* and *per concept* basis to identify extremes of scale usage and polarizing concepts. For comparative purposes, clustering solutions of a fixed number of segments (proposed: 3) will be created for 2 techniques of clustering (Agglomerative Hierarchical Clustering and K-Mean) crossed with several pre-clustering data transformation approaches. The stability of these solutions will be evaluated using discriminate analysis and nearest neighbor analysis.

Discussion: This work shows the impact of pre-clustering data transformation and clustering methods on the final consumer segmentation interpretability, the relative proportion of countries within segments, and mathematical cluster stability. Countries with extreme scale usage behaviors and products with extremely polarizing (or extremely consistent) results will be highlighted as their variability (or lack thereof) ultimately drives final clustering. We hypothesize that the balance of weight between food concept variability and country scale usage variability will be meaningfully impacted by data transformation and clustering method choices, and will subsequently impact cluster stability and interpretability. Awareness of this impact is key for sensory scientists (or their data analysis partners) when formulating global segmentation analysis strategies and final recommendations to stakeholders.

O3.4 Is Aggregate Survey Data a Misleading Representation of Individual Behavior?

Patti Wojnicz, Juan Martinez

P&K Research, Chicago, USA

We know that in society there are many factors that effect consumer behavior and cause them to change their mind on a regular basis. Hence the push for manufacturers to provide variety in order to keep consumers engaged with their brand.

However, in consumer market research we often conduct experiments in a controlled environment such as a neutral central location facility with few distractions. Within these controlled experiments we often ask the respondent to rate the same product on different measures of “appeal” (overall liking, purchase intent, preference) and/or under different contexts (unbranded evaluation followed by branded). The results are analyzed and reported in aggregate, and this approach is widely accepted in the industry. However, we rarely take the time to analyze respondent rating behavior within the test. Are the respondents consistent in their opinions? For instance, do they give stronger overall liking and/or purchase intent ratings to the product they prefer more overall? Or are they inconsistent and actually rate their preferred product lower on overall hedonics and key measures. When examining blinded vs branded ratings, we measure the “brand effect,” but do not necessarily dive into the driving factors of brand effect, or lack thereof? Is it, because respondents are consistent in their rank order of acceptance and magnitude of ratings across the product set? Or

is it because consumer opinions change across the board (e.g. dis-likers become likers once brand is known, and vice versa), and when analyzed in aggregate, similar conclusions are drawn. In particular we examine the effect that the product's place/rank within each consumer has on their shift in ratings in a blinded, then branded evaluation. As well as the effect of actual in-market brand usage. Meaning, is brand effect in a research study driven by the consumers usual brand/product usage, or does this have no/minimal impact.

P&K Research will analyze several consumer data sets to answer these questions and uncover trends in individual consumer survey rating behavior.

O3.5 Understanding Consumers by Clustering – Successes, Problems and Pitfalls.

Anne Hasted

Qi Statistics Ltd, West Malling, United Kingdom

Clustering consumers using liking scale data collected in typical consumer research projects is a widely used technique. Sometimes the products are selected with the underlying aim of identifying different groups of consumers but for some of our clients investigative clustering is carried out routinely on product sets tested.

This talk uses data from two studies, one set up to check repeatability of consumer scores and another study designed to investigate the reproducibility of clustering results. I will use these to discuss the robustness of typical clustering solutions. I also propose a simple method (available in any software) for cluster trimming, identifying and separating out respondents whose individual data is not well explained by the cluster mean.

Session 4: Modelling multiblock data sets

O4.1 The SO-PLS (sequential and orthogonalized PLS) for path modelling; method, relation to graphical modelling and applications.

Tormod Næs[1], Rosario Romano[2]

[1]Nofima AS, Ås, Norway. [2]Univ. Naples, Federico II, Napoli, Italy

In many application fields the variables used to measure a phenomenon can be grouped into homogeneous blocks that measure partial aspects of the phenomenon. For example, in sensory analysis the overall quality of a product typically depends on both the taste, odor and texture variables. In some contexts, there may exist a structure of the relations between the different blocks that gives rise to a chain of influences. The influences can be visualized using a graphical illustration usually called a path diagram. Analysing data from such diagrams are often called path modelling or structural equations modelling (SEM). The classical SEM model consists of two parts; a model for each block based on a factor analysis structure and a model linking the latent variables according to the structure of the path diagram.

In the literatures, there are two main analysis traditions. One of them is called LISREL and is based on maximum likelihood fitting of the parameters in the covariance matrix. The other tradition is called PLS path modelling (PLS-PM) and is based on optimizing covariances between latent variables. A

major drawback with these methods is that each block is assumed to be one-dimensional (one latent variable for each block), which is often not the case in practice. Remedies exist for solving these issues within the traditional frameworks, but recently a couple of alternative approaches have been proposed for solving the problem. One of these is the SO-PLS method.

This presentation is about the use of the multi-block regression method SO-PLS for path modelling. The approach splits the estimation into separate sequential orthogonalized PLS regressions (SO-PLS) for each endogenous block. The method is flexible, graphically oriented and allows for handling multidimensional blocks and diagnosing missing paths. This presentation will focus on the basic idea of the method as well as how the method relates to traditional approaches and concepts in graphical modelling such as DAGs (directed acyclic graphs) and topological order. Secondly, the paper discusses how the concept of direct and indirect effects can be defined in terms of explained variances. The method will be illustrated using sensory and consumer data.

Session 5: Analysing data from temporal and emerging methods

O5.1 A Flavor Map: Understanding Flavor Pairing Through Projective Mapping

Araceli Arellano-Covarrubias[1], Paula Varela[2], Carlos Gómez-Corona[3], Héctor Bernardo Escalona-Buendia[1]

[1]Universidad Autónoma Metropolitana, Ciudad de México, Mexico. [2]Nofima, AS, Norway. [3]Xoc Estudio, Ciudad de México, Mexico

Food and flavor pairing have been popular topics among scientists and consumer researchers who intend to find successful flavor pairings within consumers. This research was performed in order to find a methodology that could be used for this purpose. Thirty flavors, which were highly mentioned in previous studies using social media, and seven beverages were selected. All beverages (soda, white wine, red wine, tequila, mezcal, blond beer, and dark beer) were chosen due to the high popularity or growing consumption (in the case of wine) within Mexican people. Projective mapping methodology was applied to 100 Mexican consumers, in which several cards, each one with one image of each flavor and beverage, were designed. In order to create food/flavor pairing maps according to each consumer preferences, consumers were asked to position in a limited area each flavor card close to the beverage card with which it best combines and near to another flavor card that could represent a good combination. The food/flavor maps were translated to coordinates of each flavor/beverage by using fizz software®. The coordinates were analyzed through Multiple Factorial Analysis (MFA) and Generalized Procrustes Analysis (GPA), both followed by Agglomerative Hierarchical Clustering (AHC). Results showed that food/flavor pairing data could be analyzed with both methods (MFA and GPA), however, GPA describe a higher percentage of explained variance in the first two components (near to 100%), contrarily to the MFA in which a several numbers of components should be needed to reach the same values (more than 30). Additionally, AHC of GPA clustered all products (flavors and beverages) in more groups than MFA, which also provides better flavor combinations and a better interpretation of the maps. In general, fruits tend to be joined together, all vegetables were clustered together, and also, some flavors that could be paired with some beverages such as beer, wine, soda, and tequila were also clustered along with each beverage. So far, projective mapping followed by GPA showed to be also a useful tool to analyze flavor/food pairing data of consumers, however, these results remain theoretical due that consumers did not taste any products, so further research has to be done in the practice.

O5.2 Implicit and explicit validation of panelist engagement during sensory testing

Mackenzie Hannum, Christopher Simons

The Ohio State University, Columbus, USA

Engagement may drive involvement and focus during sensory and consumer testing and help generate reliable data. An Engagement Questionnaire (EQ) was recently developed to characterize consumer engagement during sensory and consumer testing with the intent to delineate between engaging and disengaging methodologies and/or panelists. Additionally, we sought to explore use of an implicit correlate—facial tracking—to fully understand in real-time what it means to be engaged, both physiologically and cognitively.

Presently, to assess performance of the EQ, the instrument underwent a known-groups analysis and was evaluated after two conditions: a more engaging and a less engaging sensory task. For the disengaging task, subjects (N=8) participated in a discrimination test consisting of 20 tetrads. The samples were the same sucrose concentration to ensure a state of helplessness and encourage disengagement with the task. For the engaging task, subjects (N=7) participated in 20 tetrad tests, however, the tetrads spanned a range of stimuli and difficulty. To instill internal motivation, subjects received immediate feedback after each tetrad. Facial expressions and emotion data were collected throughout the experiments, using automatic facial emotion recognition software (FaceReader, Noldus Technology, Netherlands) that tracks emotional valence and arousal on the circumplex model. After participating in the task, subjects rated their level of agreement on the EQ.

The EQ successfully discriminated between the disengaged and engaged conditions. Subjects were more actively involved in the task ($t=3.001$, $p=0.047$), experienced higher levels of purpose and intent ($t=2.623$, $p=0.021$) and more affective value ($t=2.642$, $p=0.020$) when in the engaged condition compared to the disengaged condition. When averaged across all 20 tetrads, subjects experienced increased valence levels in the engaged condition ($t=8.083$, $p<0.001$); arousal was similar in both conditions. Furthermore, approximately half-way through the experiment, valence levels for the two groups significantly diverged, with panelists experiencing higher levels of valence in the engaged condition for the duration of the task. Results suggest an engaging test design can maintain and sustain levels of engagement for a longer duration.

Now validated via known groups analysis, the EQ can confidently be used to help differentiate varying levels of engagement. The outcome of this research could enable predictive modeling amongst implicit correlates of engagement and data quality. Such an instrument allows for a deeper understanding of panelist engagement and its impact on data quality, enabling the design of sensory tests capable of generating more reliable data both within the field of food science and beyond.

O5.3 Identifying Temporal Drivers Based On Temporal Sensory Descriptions And Overall Consumer Expectations

Quoc Cuong Nguyen[1], Paula Varela[2]

[1]Department of Food Technology, Ho Chi Minh city University of Technology, VNU - HCM, Ho Chi Minh City, Vietnam. [2]Nofima, Ås, Norway

Capturing temporal sensory changes has been considered in recent research to better understand how consumers perceive food products. This information can be linked to consumer expectations

(e.g., liking, satiety) to study the drivers throughout the eating experience, namely temporal drivers. It is assumed that consumers have different strategies to manage food in mouth; these individual differences should play a role both on consumer perception of the products and their expectations of liking or satiety. This study explores the use of penalty-lift analyses for each time point to identify the temporal drivers of liking/ satiety for different groups of consumers, after segmentation.

Eight yoghurt samples, based on DOE, with identical composition, varying in textural properties, were used in the study. Temporal Check-All-That-Apply (TCATA) was used to describe dynamic sensory profile. Consumers (n=101) tasted each sample and rated their liking and expected satiety.

Cluster analysis of variables around latent variables (CLV) method was applied to cluster consumers based on their expectations of satiety. There were two main clusters: cluster 1 (n = 36), cluster 2 (n = 58), and a noise cluster (n = 7). From that, cluster 1 and cluster 2 were used to identify temporal drivers of liking and satiety.

Penalty-lift analyses were used for each time point. Also, the false discovery rate (FDR) was applied to correct p-values for multiple tests responding to sequential time points. Both clusters presented similar drivers linked to the textural properties of yoghurt. However, differences were highlighted related to the particle size attributes and flavour intensities. For cluster 1, while *Gritty* was positive driver from the middle to the end, *Sandy* as negative driver at the middle; *Vanilla* was positive driver during the mastication. For cluster 2, only *Sweet* was pointed as positive driver at the beginning, and *Dry* as negative driver in some time points at the middle of the mastication.

With regard to expected satiety, the only difference was that *Gritty*/*Sandy* was considered as positive/negative driver for cluster 1, but not for cluster 2; they were all significant over time.

These findings demonstrated that the temporal driver approach is a suitable method to unveil the drivers of liking/satiety during the eating process in segmented consumers.

O5.4 Statistical inference for TDS data modeled by Semi-Markov processes

Cindy Frascolla[1], Guillaume Lecuelle[2], Hervé Cardot[1], Michel Visalli[2], Pascal Schlich[2]

[1]Institut de Mathématiques de Bourgogne, UMR CNRS 5584, Université de Bourgogne, Dijon, France. [2]Centre des Sciences du Goût et de l'Alimentation, UMR AgroSup Dijon-CNRS-INRA-Université de Bourgogne, Dijon, France

Lecuelle et al. (2018) proposed modeling Temporal Dominance of Sensations (TDS) data thanks to semi-Markov chains in order to take into account the transitions from one descriptor to another, which classical methods never acknowledge, as well as the dominance duration law of each attribute. In the same stochastic framework, the current paper proposes a statistical test for comparing the temporal perception of two products. Classical TDS analysis either compares products time by time thanks to the curves of differences of dominance rates (Pineau et al., 2009), but this method essentially relies on a visual investigation, or based on the durations of dominance by ANOVA and MANOVA, which again ignores transitions. The proposed statistical test is based on the likelihood ratio test between two semi-Markov models. Three approaches are evaluated to build the critical region: a first one based on the parametric bootstrap, a second one on permutations and a third one on the asymptotic law of the likelihood ratio. Nominal level and power of these three

approaches are compared thanks to Monte-Carlo simulations. It turned out that the best approach is the permutation test.

Once a statistical difference between two products has been established, it is of interest to find out whether the difference comes from transition probabilities, from duration of sojourn times or from both. Specific likelihood ratio tests are proposed for answering this question, which are also based on the permutation approach.

This new methodology is applied to several real datasets and the outcomes are compared with those from classical TDS analysis (TDS difference curves and CVA). The paper highlights when and why the results from this new approach agrees with those from the classical ones or not.

O5.5 On statistical methods for TDS data analysis: Consideration about characteristics of each panelist and each taste

Sumito Kurata[1], Reiko Kuroda[2], Fumiyasu Komaki[1,3]

[1]The University of Tokyo, Tokyo, Japan. [2]LOTTE CO., LTD. Central Laboratory, Saitama, Japan. [3]RIKEN Center for Brain Science, Saitama, Japan

In the sensory evaluation, TDS (Temporal Dominance of Sensation) data analysis is one Representative method. TDS data is time series that consist of the temporal change the most "dominant" taste that each panelist (test subject) selected. The candidates of taste (e.g., "Sweet", "Bitter", "Salty", "Fruity",and so on) are given in advance.

Up to now, many studies analysed TDS data by drawing "TDS curve" that shows the change of dominance rates based on the answers of panelists. Although TDS curve is visually understandable, this method cannot consider the individual variation fully because it is regarded as analysing only the mean of the data. Thus, there is concern that minority opinions might ignore, moreover, this method will not cope well if the data have "branch" (for example, some panelists felt "Sweet" after "Bitter", and other ones felt "Fruity" after "Bitter").

In recent years, some statistical methods are being introduced (e.g., Franczak et al. (2015)). One of them is the Markov model (MM): a stochastic model under the assumption that each state depends on previous states, not future ones. This method can adopt each opinion, and it can express the branch of the data by using the transition probability. Lecuelle et al. (2018) applied the semi-Markov model (SMM) to the TDS data. SMM is an extended model of MM, it can consider the wide class of probability distribution of the sojourn time. In the TDS data, the "sojourn time" corresponds to the time length of one taste dominates the panelist's sense.

We consider that there is plenty of scope for improvement in the analysis of the TDS data based on MM. Particulary, we focus on the characteristic of the "panelists" and "taste". Each panelist has his or her own sense and preference, also each taste (e.g., "Sweet" and "Bitter") has its feature. Therefore, without consideration about them, we cannot gain the valid conclusion from the TDS data. In this presentation, We introduce some new approaches that are adapted to the TDS data, and we show an example of real data analysis.

Session 6: Getting more out of sensory and consumer data

O6.1 Men, masculinity, and flavors: A multidimensional social representation

Carlos Gómez-Corona
Firmenich SA, Mexico City, Mexico

A social representation is a set of elements that are functionally articulated between them. These elements can be a set of concepts, ideas, images, attitudes, etc. The representation system has three dimensions: a set of information which are the knowledge that individuals have towards an object, an attitude that marks the positive-negative connotation of an individual or group towards the object; and a field of representation, that means a structure that organizes the elements of the representation.

There are many ways to access or elicit a latent social representation, the most common being a free word association task. The analysis usually relies in the intersection of frequency of elicitation versus importance of words. Such an intersection can depict a structure and dimensional-based on similarity and importance but does not take into consideration the valence of each element. This may be acceptable for understanding social representations, but for those that involve consumer behavior or choice, it's vital to be able to include additional information such as their valance. In other words, the positive or negative aspects of the representation.

To better understand the organization of the space of a social representation between flavors and an abstract social construct such as masculinity, the author conducted a study with 240 men consumers in Mexico using a free word association task to understand the representation of "man", "masculinity", and "masculine flavors". The task was followed by a ranking, and the valence of each word was measured in a scale from -1 (negative connotation) to +1 (positive connotation). A dissimilarity matrix of the data was submitted to a multidimensional scaling (MDS), one for each inductor word. The results show that is possible to identify how the space of the representation is defined by each word. For example, in the case of "man representation" the space goes from "courage" to "masculine" in Dim1, and from "male" to "father" in Dim2. The MDS map with a Kruskal's stress of 0.130, shows the relationship between the words, e.g. "manly" related to "sex" and "formal". This multidimensional approach was also used to explore the representation of masculinity and masculine flavors, with similar results. We conclude that this approach closes the gap of the understanding of the space of a social representation and enables a new understanding of the organization of each element of the representation, which has not yet been previously explored by the social representation theory.

O6.4 Going deeper in the analysis of contingency tables: application to Check-All-That-Apply and Free-Comment data

Benjamin Mahieu[1,2], Michel Visalli[1], Pascal Schlich[1]

[1]Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, Univ. Bourgogne Franche-Comté, Dijon, France. [2]SensoStat, Dijon, France

Check-All-That-Apply (CATA) and Free-Comment (FC) provide a so-called contingency table containing citation counts of words or descriptors (columns) by products (lines). This table is most often analysed using correspondence analysis (CA). CA is based on chi-square distance and aims at decomposing dependency between products and descriptors into axes of maximal and decreasing dependencies. This makes sense if the dependency is first established by a chi-square test. However, the p-value of this test is not valid when the contingency table contains too many low expected citation rates, a situation likely to occur, especially with FC data. To overcome this problem, a Monte-Carlo approach is proposed to compute an empirical p-value by comparing observed chi-square statistic to the distribution of those simulated under null hypothesis of independence. A significant chi-square test only means that at least the first CA axis captures some dependencies. From that, there is a need to know how many other axes are also relevant. Based on the same set of simulated contingency tables, a succession of adapted chi-square tests is proposed to test dependence of each axis successively. CA of a contingency table ignores the panellists' variability and provides no confidence ellipses around products. The paper proposes to compute confidence ellipses thanks to total bootstrap in which Procrustes rotations are performed by considering the space earlier defined by the number of relevant axes. Finally, sensory interpretation requires indicating which descriptors are statistically more or less cited for each product. Chi-square per cell is sometimes used for that purpose, but chi-square distribution is often not valid. The exact Fisher test should rather be used. Still these tests consider the full space, while we want to rely only on the relevant axes. We propose to "reverse" the CA algorithm to determine, from the coordinates of products and descriptors on the relevant axes, the theoretical contingency table on which exact Fisher tests can thus be conducted. The usefulness of these new statistical tools will be demonstrated on real CATA and FC data.

O6.5 CATARACT, a new procedure to evaluate the quality of CATA data

Thierry Fahmy[1], Fabien Llobell[1,2], Amaury Labenne[1]

[1]Addinsoft, Paris, France. [2]ONIRIS, Nantes, France

CATA (Check-All-That-Apply) questionnaires present several advantages for the individuals being surveyed. Given that the choice is limited to two alternatives, they reduce the complexity of the decision making for choosing a response. The decision process is quicker, easier, but as the effort is less, the quality might also be worse. As this type of surveys is often conducted on consumers that are not trained, the risk of obtaining noisy data is higher than with surveys with more than two possible answers ran on trained panels.

When our team worked on developing new methods for describing and visualizing general survey data, special attention was paid to the specific case of CATA data, with a benefit for sensory analyses. The CATARACT (CATA Rejection and Acceptation Tests) procedure allows to give a description of CATA surveys and to extract information on the quality of the questionnaires in several successive phases. Self-explanatory visualizations are provided as part of the procedure to facilitate the communication of the results. The CATARACT procedure could be seen as a first step and complement to the procedures developed by Meyners et al. (2013, 2014) and to the CATATIS procedure recently developed by Llobell et al. (2019).

The descriptive analysis that is provided is based on techniques and tactics that are aimed at helping the analyst to identify potentially suspicious observations. The tests that are run are independence exact tests that had not been developed so far and that are validated and will soon be submitted to a journal (Fahmy, 2020). The sensory sciences community will benefit from these new developments in 2020 through the XLSTAT Sensory software developed by Addinsoft.

Addinsoft (2019). XLSTAT statistical and data analysis software. Paris, France. https://www.xlstat.com.

Fahmy T. (to be submitted in 2020). A new algorithm for a multidimensional version of the Fisher's exact test.
Llobell F, Cariou V., Vigneau E., Labenne A. and Qannari E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. Food Quality and Preference, 72, 31-39.
Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. Food Quality and Preference, 30(2), 309–319.
Meyners, M., & Castura, J. C. (2014). Check-all-that-apply questions. Novel techniques in sensory characterization and consumer profiling (pp. 271–305). CRC Press.

Flash presentations 6th October

F1.1 Can children use temporal check-all-that-apply (TCATA) and temporal dominance of sensations (TDS)?

Ana Laura Velázquez[1], Leticia Vidal[1], Paula Varela[2], Gastón Ares[1]

[1]Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Pando, Canelones, Uruguay. [2]Nofima, Ås, Norway

One of the strategies that can be implemented to promote healthier eating patterns among children is the development of healthy products that meet their sensory and hedonic expectations. Since their needs and wants differ from those of adults, involving them in the product development process is essential. In the recent years, sensory scientists have adapted several sensory methods to fit children's cognitive abilities according to the different developmental stages. Although children have been reported to being able to use sensory methods to describe foods and beverages, applications are limited to static characterizations. In this context, the objective of this study was to evaluate the suitability of two dynamic methods for sensory analysis with children: temporal check-all-that-apply (TCATA) and Temporal dominance of sensations (TDS). A video featuring colored circles was used to convey the idea of temporal changes without food cues while a series of six vanilla milk desserts were used as case study. A total of 102 children (8 to 12 years old) recruited from two private Uruguayan schools participated in the study. They were randomly divided in two groups: one group used TCATA to describe the video and the vanilla milk desserts, while the other group used TDS. Results showed that both methods allowed capturing the dynamics of the video. However, TCATA provided a more detailed description of how the colored circles evolved with time than TDS. In the case of the milk desserts samples, both methodologies showed similar results regarding the most relevant sensory characteristics but children mostly used them as static methods. The concept of dominance seemed to be particularly complex for children. In the TDS task, children dithered for a long time before selecting a color or an attribute, which lead to significantly longer starting times compared to the TCATA task. In summary, results suggest that refinements are needed to make the methods applicable with children.

F1.2 Analysing time intervals of TCATA citation rates using linear mixed model ANOVA

Jonas Yde Junge[1,2], Line Ahm Mielby[1], Yan Zeng[3], Yuanxia Sun[3], Derek V. Byrne[1], John C. Castura[4], Ulla Kidmose[1]

[1]Food Quality Perception & Society, Department of Food Science, Aarhus University, Aarhus, Denmark. [2]Sino-Danish Center for Education and Research, University of Chinese Academy of Sciences, Beijing, China. [3]Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China. [4]Compusense Inc., Guelph, Canada

The quest to find palatable ways to reduce the sugar content of foods and beverages has focused mainly on the intensity of sweetness, the side tastes of non-caloric sweeteners, and on the temporal profiles of these sweeteners. Previously, Temporal Check-All-That-Apply (TCATA) and other methods have been used to investigate the temporal profiles of sucrose and non-caloric sweeteners. Data have been analysed with multivariate methods suitable for exploratory investigations. Here, we apply ANOVA for hypothesis testing using TCATA data.

The aim of the study is to investigate the temporal taste differences between different sweet solutions, and how these differences are affected by an addition of citric acid. Further, the study aims at investigating the relationship between temporal profiles and intensity ratings.

We investigated the temporal profiles of three different equi-sweet solutions (sucrose, D-allulose, D-allulose + mogroside), with and without citric acid using both a) intensity ratings, and b) TCATA measures. The study was performed with >100 Chinese consumers.

By summing the citations for a given time interval, a measure of fractional “Area Under the Curve” (AUC) is constructed for the specified time intervals. These AUC can then be compared using ANOVA based on linear mixed models. Thereby we are able to test hypotheses regarding the differences resulting from the different tastants.

F1.3 Common problems in the experimental design of sensory tests in agricultural studies and recommended solutions

Masoumeh Bejaei

Agriculture and Agri-Food Canada, Summerland, Canada

Sensory research projects relevant to raw agricultural products (especially fruits and vegetables) can be designed using various research methods. There are many factors, some unique to the products, that should be considered in the design of those experiments. Most of the raw agricultural products are perishable foods, produced under natural conditions, and researchers have little or no control over many production factors that may influence the quality of the products or their sensory attributes. Finding exact replicates in raw agricultural products is not feasible in most cases because between cultivar and within cultivar variations exist among them. Variations even exists within one fruit/vegetable; an attribute can differ from top-to-bottom and side-to-side of the fruit (e.g., sun-exposed side or shaded side). In some agricultural studies, samples were prepared without

considering these variation sources. For example, if slices of fruits were provided to judges in a sensory test, those slices were not excised from specific parts of the fruits (e.g., the sun/shade transition zones in apples) to reduce the extra variation sources. In addition, the variations within experimental units required provision of multiple biological replicates from the same treatments for each judge (to capture random variation and evaluate the judge performance), and repetition of the study to collect an independent validation data set to be able to generalize the results.

In some studies, the applied tests did not match the scales of the measurements as a result statistical test assumptions were not met. One of the common observed problems in agricultural studies has been related to the design of tests in which sensory ratings were conducted by the same judges multiple times in an experiment (e.g., in a storage study) nonetheless the data were analyzed without the use of the repeated measures models. For example, they analyzed treatment and time effects using independent tests or in a two-way ANOVA without meeting the independent-measurements assumption of parametric tests. Identifying whether time was a fixed effect or a random effect has not been done in a clear and consistent way in some cases. There were also studies in which a treatment effect was considered as a between-subject factor even though all judges tasted samples from all treatments.

Several examples will be presented to demonstrate the impact of the common research design and statistical analysis problems in studying the sensory characteristics of raw agricultural products and solutions will be discussed.

F1.4 Do we look at chocolate labels differently when deciding on purchasing, liking or healthiness? An eye-tracking study.

L. Porras[1], R. Silva[2], L. Laguna[3], P. Puerta[3], A. Tarrega[3]

[1]Universidad de Antioquia, Medellín, Colombia. [2]Universidad Peruana Unión, Lima, Peru. [3]IATA-CSIC, Valencia, Spain

This study examines the differences in attention paid by consumers to the different elements on the dark chocolate labels when deciding on purchasing, liking or healthiness. Three different slides with four images of dark chocolate packages were constructed using products of different brands. Each slide included four types of chocolate; 1) basic dark chocolate; 2) chocolate with high percentage of cocoa; 3) high percentage of cocoa and sugar reduced chocolate and 4) high percentage of cocoa, sugar reduced and with stevia chocolate. One hundred-twenty chocolate consumers (69% female and 31% male; 21 - 59 years old) participated in the study. During the evaluation session, a first slide was presented to the participant and they would choose the chocolate that they would buy, then a second different slide was presented where the choice of chocolate that he / she expected to like the most was made. Finally, the participant was asked to choose the healthiest chocolate from a third different slide. The slide presented in each question (purchasing, liking or healthiness) and the position of the four chocolates on each slide was varied among consumers following a balanced design. During the session, participant's eye movements were recorded using an eye-tracking device (Pupil Labs GmbH, Berlin, Germany). The percentage of consumers that looked at each relevant element on the slide (brand names, % cocoa, sugar reduction claim, stevia logo, ingredients, nutritional facts, and chocolate image) and the number of fixations on each element were calculated.

On observing the percentage of consumers that paid attention to the elements, differences were found among the questions. Elements that received attention from most of the consumers (70 - 80%) were the chocolate image and nutritional information for purchasing; chocolate image and the "dark chocolate" term for liking; and the nutritional information, chocolate image, and list of ingredients for healthiness. The number of fixations significantly varied among questions for most of the elements. Percentage of cocoa, sugar reduced, the term "dark chocolate," and chocolate image presented higher numbers of fixation when deciding on liking and purchasing than when deciding on healthiness. On the other hand, the nutritional information and ingredients received a higher number of fixations when deciding on purchasing and healthiness than when deciding acceptability.

Flash presentations 7th October

F2.2 Connecting Trained-Panel Degree of Difference to Other Discrimination Methods

Katie Osdoba, Yvonne Koelliker, Annlyse Retiveau Krogmann

Sensory Spectrum, Inc., New Providence, USA

The objective of many sensory evaluation studies is to quantify differences between two or more products and understand if these differences are meaningful in a real-world context. This is often achieved using descriptive analysis (by looking at statistical differences among products in attribute intensities) or discrimination testing methods (by comparing d-prime estimates to pre-established criteria). Sensory Spectrum's highly-trained descriptive panels regularly use Degree of Difference (DOD) ratings as a proxy for full descriptive profiling and/or traditional discrimination testing. This research serves to augment our Sensometrics 2014 poster comparing DOD to triangle testing. The primary objective is to understand the link between DOD ratings given by a trained Spectrum Descriptive Analysis panel, full descriptive analysis results, and results from more traditional choice-based discrimination methods (tetrad, 2-AFC, etc.). To achieve this objective, data are collected from studies in which trained-panel DOD ratings plus one or more other methods of objective sensory evaluation are carried out on the same set of products. The panel's DOD ratings are compared to d-prime estimates (from traditional discrimination methods) or statistically significant differences (from full descriptive profiling). Results show that trained-panel DOD ratings lead to the same overall conclusions as traditional descriptive analysis and discrimination methods. Results from the various methods are further compared to better understand the link between the methods and help practitioners select appropriate methods for their objectives given their resource constraints. For sensory programs with well-established descriptive panels, training them to rate degree of difference provides a quick and simple method for approximating results of discrimination testing or full descriptive profiling. The results of these trained-panel DOD evaluations may serve as stand-alone findings, aid in interpretation of descriptive analysis results, or assist product screening/selection for more extensive research projects. As an added benefit, these trained panels can also rate DOD on one modality at a time (e.g. if there are appearance differences between products, but only flavor differences are of interest) and/or provide qualitative, directional comments on product characteristics to aid product developers or quality partners.

F2.3 Taste versus reputation: the impact of sensory evaluation on consumers' preferences for pomegranate arils

Alice Stiletto, Erika Rozzanigo, Samuele Trestini

University of Padova, Padova, Italy

In recent years pomegranate (*Punica granatum L.*) gained more attention from both producers and consumers due to its health benefits. Despite this, from the literature analysis it emerges that only few studies aim to link consumer preferences and product characterization to explain food choices. In addition, research interest has been focusing mainly on juice, while little is known about the fresh fruit. In this context, the purpose of this research is to combine sensory analysis and choice experiment (CE) to evaluate the role of tasting on consumers' purchasing choice and on the willingness to pay (WTP) for different attributes. Three samples of arils, derived from fresh pomegranates of different cultivars and with different origin, have been evaluated through a consumer test on 400 people. Considering that pomegranate is not a well-known fruit and that many consumers have not had previous experience of consumption, half of respondents underwent a sensory test before the choice experiment (CE) survey, for which 100g of packaged arils have been proposed. The remaining consumers participated to the CE session, supported only by a visual evaluation of the products, simulating a typical free-service purchasing process. Questionnaires were administered in Veneto region (Italy) through a face-to-face interview in November 2019. The three samples have been also characterised by chemical analyses and a panel test. Results are expected to be coherent with the literature concerning the relevant role of hedonic evaluation on consumers' preferences. We expect that sweeter cultivars and those with red arils are going to be the preferred ones, while high levels of acidity and seed intrusiveness are going to be considered negative elements. Beside this, we suppose that low degrees of bitterness and astringency are going to be preferred. Furthermore, we believe that, being pomegranate an innovative product, when consumers have the opportunity to taste the arils, the value associated to other product cues (i.e. reputational ones like country of origin and place of purchase) may change if compared to a mere visual evaluation. In this way, we expect to contribute to the current literature with some innovative results, comparable with those raised from studies in other sectors (e.g. wine). Finally, the link with chemical and panel data could allow to improve the ability to forecast both consumers' liking and WTP and suggest marketing strategies to communicate sensory traits.

F2.4 Sensory Detection of Wine Faults Over Time Using Flash Profiling and the Electronic Tongue

Victoria Paup, Tara Cook, Charles Edwards, Carolyn Ross

Washington State University, Pullman, USA

Wine faults are negative sensory attributes in wine that may result in large economic losses for wineries. One cause of wine faults is the growth of spoilage microorganisms. As these microorganisms can grow in-bottle during storage, rapid detection methods are needed to allow for early remediation. Since identifying quality changes quickly is critical, the use of a rapid sensory profiling method, such as flash profiling, is very useful. Thus, the objective of this study was to use the electronic tongue (e-tongue) and a rapid sensory method, flash sensory profiling, to evaluate

changes in wine over time due to the presence of different spoilage microorganisms. Merlot wine was bottled and inoculated with one of the following common wine spoilage microorganisms, including two strains of *Brettanomyces*, *Lactobacillus brevis, Pediococcus parvulus,* and *Acetobacter pastueriansus.* These organisms are associated with sensory attributes such as barnyard, vinegar, buttery and geranium. Starting at Day 0, wines were analyzed weekly until Day 42 using the e-tongue and flash profiling. The growth of the microorganisms was monitored through microbial plate counts. For flash profiling, experienced panelists (n=7) selected attributes that differentiated the samples and ranked all samples for these attributes. Principal Component Analysis, Generalized Procrustes Analysis, and Agglomerative Hierarchical Clustering analyses were performed. Over 42 days of storage, both flash profiling and e-tongue analyses differentiated the wines. Microbiological data reflected growth of the different microorganisms, with plate counts increasing over time for *Brettanomyces* and *Pediococcus*. The e-tongue displayed low discrimination among the different microorganisms through Day 14 of storage. However, at Day 21, the discrimination index of the e-tongue wines containing different microorganisms increased to 91%, and this high discrimination continued until Day 42 of storage. From the flash profiling data, an increase in citation of sensory attributes associated with faulted wines was noted over time, with specific attributes including barnyard and medicinal flavors and aromas. These citations reached a peak at Day 42 of microbial growth, when 18 out of 31 attributes used to describe differences among the wines were related to wine faults. As the e-tongue detected differences starting at Day 21, but sensory differences were not apparent until Day 42, these results suggest that the e-tongue is a useful tool for early detection of wine faults and possible remediation. The application of these novel techniques (e-tongue partnered with flash profiling) may be the key to detecting and limiting financial losses associated with wine faults.

F2.5 Consensual model of data processing based on the blockchain approach for sensory data.

Vladimir Vietoris, Patricia Martisova, Patricia Mackova, Lucia Sekanova

Slovak University of Agriculture, Nitra, Slovakia

In our scientific work, we tried to apply the consensus model adopted from blockchain techniques to sensory methodology. The general innovation of the data processing is not calculating the mean values of the panel (products/weights of the assessors) but a few alternative approaches to the consensual evaluation of the data. We made experimental work on symmetric (equal number of products and assessors) and non-symmetric (not equal number of products and assessors) sensory matrices. Results were compared to "classical" application of the scales afterwards. Findings will be presented.

Workshop 1:

Applying Text Mining Methods for Sensory Evaluation Research

Organizers: Sébastien Lê, Jacob Lahne
Participants: Sébastien Lê, Anne Hasted, Jacob Lahne, Alexiane Luc, Benjamin Mahieu, Leticia Vidal

Facilitator: Sébastien Lê, Anne Hasted

Jacob Lahne, Web scraping for sensory research: a case study with cocktails.

Websites that review or describe food products are potentially rich sources of sensory data. However, many sensory scientists find acquiring and parsing web data a challenging barrier. This talk introduces web scraping for sensory data using a case study of a website with ~4500 well-structured cocktail recipes. A workflow for acquiring and parsing the data is presented (in R) with a focus on exposing the basics of interacting with HTML trees (using R). The results are analysed using featural (network) approaches to make the sensory aspects of the data visible.

Benjamin Mahieu, Sensory characterization of home-perfumes using Free-Comment as response to open-ended questions.

This talk explores text data collected using a Free-Comment protocol in which 88 consumers evaluated 4 home-perfumes by answering the question “Describe the olfactory characteristics of this perfume”. The talk presents data pre-processing (cleaning, lemmatization, etc.) and proposes an original approach based on a classification using chi-square distance to group descriptors that are “close enough” to be associated. Outputs of the statistical analyses of the pre-processed data are presented to show the relevance of the Free-Comment methodology in discriminating and characterizing a set of products.

Alexiane Luc, *NLP strategies to analyse consumer data using valency: an application to Free JAR data*. This talk presents the analysis of consumer free comment with a constraint of expression, which appears in the systematic use of a JAR structure to describe the tested products. This type of data is also known as Free JAR data. A new algorithm which considers all the specificities of Free JAR consumer data to understand the opinion generated by the tested products among the subjects, will be presented.

Leticia Vidal, *Automatic text analysis of Twitter data*. Social media is a valuable source to study consumer behaviour. Some years ago, ~50,000 tweets containing the words “breakfast”, “lunch”, “snack” or “dinner” were retrieved to investigate “what people say when tweeting about eating situations”. Content analysis was used, but due to the time-consuming nature of manual coding it was only applied to a subset of 16,000 tweets. In this talk, the use of automatic text analysis tools to gain valuable insights based on the whole dataset will be explored.

Workshop 2:

Joint SSP and sensometrics WORKSHOP: Artificial Intelligence in Sensory Practice: Separating Promise from Hype

Lead by Rafal Drabek
Participants: Amanda Grzeda, John Ennis, Leah Hamilton

The rapid evolution of computational technology has sparked excitement in the ability of artificial intelligence (AI) to provide never-before-considered solutions and insights. The nebulous nature of the term "AI," along with the deep technical knowledge AI seemingly requires, has resulted in high expectations – what *can't* AI do? And, while most agree the true power of AI will be unleashed when humans and machines work seamlessly together - each leveraging the strengths of the other - how will that goal be realized in sensory practice? This workshop separates the promise from the hype.

In Part 1, **Amanda Grzeda** shares within-business experiences. First, how to effectively translate AI excitement into strong business questions. Next, how to source the right domain knowledge, which likely lies across multiple individuals who represent different departments and functions. Then, how to cast the net wide enough with regard to metadata - considering what's relevant today, and what might be relevant in the future. Finally, with a model in-hand, how to best play the role of the human in the human-machine interface. Data modeling is not new to us, but given the focused attention, our ability to communicate it effectively must evolve.

In Part 2, **John Ennis** discusses how, whenever a new general-purpose technology - such as steam, electricity, digital computing, or AI - appears, industrialists must decide how to transform. One option, "fast caterpillar," is to continue in one's present activities, leveraging the new technology to increase output. The second option, "beautiful butterfly," is to use the new technology to produce results previously unattainable. In this part, Dr. Ennis reviews how AI makes both options available to sensory scientists and provides recommendations on how to balance these options to provide business benefits while effectively preparing for the future.

In Part 3, **Leah Hamilton** discusses opportunities and challenges in incorporating existing AI technologies into novel solutions for sensory science. Language-standardization training or analysis is often the bottleneck in sensory research, making the rapid analysis of thousands of words of natural language using AI an exciting prospect. There are great opportunities and challenges in using existing tools for novel sensory applications, as demonstrated through a case study that used descriptors from 6,598 reviews of international whiskies to create a flavor wheel. This talk highlights the power of using domain expertise to adapt existing AI, rather than reinventing the wheel.

The workshop concludes with panel discussion and open Q&A.

Virtual Posters Sensometrics 2020

P1 How to Conduct Business Relevant Difference Testing

John Ennis[1], Gemma Hodgson[2]

[1]Aigora, Richmond, VA, USA. [2]Qi Statistics Ltd., Berkshire, United Kingdom

How can sensory and consumer scientists incorporate the possible business consequences associated with their decisions to accept or reject samples into their analysis of sensory difference testing data? In this presentation, we first explain why classical approaches based on p-values are insufficient to answer this question and can even be misleading. Instead, we propose a solution comprised of Bayesian estimation and Thurstonian modeling. The first of these tools, Bayesian estimation, allows us to assess the costs associated with various errors in estimation, while the second, Thurstonian modeling, provides a framework for a psychologically-informed interpretation of results. Combining these two tools, we show how it is possible to quantify our beliefs about how likely a sensory difference is to be consumer-relevant and to frame our knowledge in terms of business risks. We conclude by answering our original question, showing how quantified beliefs can be combined with knowledge of possible business consequences to provide well-informed recommendations.

P2 How Linear Programming Solves a Surprising Variety of Consumer Research Problems

John Ennis[1], William Russ[2], Francis Rossi[3]
[1]Aigora, Richmond, USA. [2]The Institute for Perception, Richmond, USA. [3]PepsiCo Research and Development Data Science and Analytics, Plano, USA

The increasing ease of data collection has driven a dramatic increase in the number and size of datasets available to consumer researchers. Fortunately, a surprising variety of consumer research problems – such as TURF analysis or portfolio optimization following conjoint or MaxDiff research – can be reformulated as linear problems, opening the door to high-performance analytic techniques such as linear programming. In this presentation, we illustrate this point through examples. We start by explaining how the problems listed above can be recast as suitable for linear programming. Since linear programming solvers are freely available, this portion of the presentation alone provides significant value to practitioners. We then explain how the flexibility of linear programming allows for generalizations of common techniques. We conclude by considering the novel, emergent properties that arise when the speed of linear programming is exploited to consider such large problems as analyzing the space of all TURF solutions for a single survey.

Notes: Portions of this research were conducted while the first author was at The Institute for Perception. The views expressed in this abstract are those of the authors and do not necessarily reflect the position or policy of PepsiCo, Inc.

P3 Individual repeatability in projective mapping: use of hierarchical cluster analysis

Marcelo Miraballes, Natalia Hodos, Gonzalo Da Rosa, Adriana Gámbaro
Universidad de la República, Montevideo, Uruguay

The global repeatability of the results obtained using projective mapping (PM) is usually analysed by including blind duplicated samples on the sample set and then comparing their positions on the product spaces obtained.

The objective of this work was to explore the evaluation of the individual repeatability on projective mapping data by using hierarchical cluster analysis (HCA). Four different projective mapping data were analysed: drinkable yoghurts (two sessions, 13 assessors) spreadable processed cheese (12 assessors); and orange-flavoured drinks (13 assessors). In all PM tasks, a blind duplicated sample was included. Different semi-trained assessor panels carried out all of the PM sessions.

HCA on the individual coordinates of the PM data was performed using Euclidean distances and Ward's linkage to evaluate the performance of each assessor. Individual data were considered reliable if the duplicated samples were located in the same group, according to HCA.

Multiple factorial analysis (MFA) was carried out considering the coordinates of each assessor as a different group of variables. Confidence ellipses (95%) were calculated using truncated total bootstrapping.

For each dataset, two MFA were obtained: one considering all of the assessors' data (MFA1) and the other considering only data from reliable assessors (MFA2) according to HCA.

The data were analysed using XL-Stat software and R software. Results showed that using HCA on individual data helped identify unreliable assessors. When comparing the results of MFA1 and MFA2 of all data sets, it found that if unreliable assessors where not taken into account in the analysis, confidence ellipses were lower and less overlapping, increasing discrimination among samples. This pattern was observed on the results of all datasets analysed. Moreover, the explained variance was higher on the results of MFA2 than MFA1 for all datasets analysed.

In conclusion, the inclusion of blind duplicated samples on PM and the use of HCA to analyse individual repeatability might be a useful complement to improve the results obtained by the use of PM with semi-trained assessors.

P4 Panel and Panelist Performance Strategy in Discrimination Testing

Katie Osdoba, Annlyse Retiveau Krogmann, Lisa Codespot
Sensory Spectrum, Inc., New Providence, USA

Discrimination panels are widely used in the CPG industry for a variety of research objectives. They are used for routine assessments in product maintenance, quality control, and shelf life, as well as for product improvement and innovation initiatives. Discrimination programs work best when the project team has deep understanding of 1) consumer-relevant criteria for test design (often based on historical knowledge of products and current business imperatives) and 2) panel capability and overall performance. Few resources may currently be found in the literature on how to assess a discrimination panel's competence in terms of sensitivity, accuracy, and reliability. The primary objective of this research is to develop and document strategy and key performance indicators for discrimination panels that participate in a variety of test types – overall- and attribute-specific tests, difference and similarity tests. These indicators include techniques to assess:

- *Overall panel performance*
    1. Repeatability: When the panel completes replicates of the same set of products, are the same results achieved? Conclusions drawn?
    2. Discriminability: Does the panel adequately differentiate products that are expected to be different?
    3. Validity: Does the panel give results that are aligned with expectations based on prior knowledge or other data types?
- *Individual panelist performance*
    1. Repeatability: Can the individual panelist give the same results across multiple replications?
    2. Discriminability: When the overall panel successfully differentiates products, can the individual panelist as well?
    3. Validity: How well is the individual panelist aligned with the overall panel results?

The techniques described in this research may be used by sensory professionals to monitor their discrimination panels to ensure high performance. Overall panel performance assessment leads to greater understanding of the capabilities of the discrimination testing program and results in high confidence in the panel results for personnel relying on the data generated. Individual panelist

performance metrics may be used to identify underperforming panelists and inform remediation strategies. Implementation of both panel- and panelist-performance strategy leads to high-performing panels that provide sensitive, powerful, and reliable test results.

P5 Segmenting Consumers According To Both Consumer Perception And Preferences Using 2-Step Clustering Approach

Quoc Cuong Nguyen[1], Ingunn Berget[2], Giorgio Russolillo[3], Paula Varela[2]
[1]Ho Chi Minh City University of Technology, VNU-HCM, Ho Chi Minh City, Vietnam. [2]Nofima AS, Ås, Norway. [3]Conservatoire National des Arts et Métiers, Paris, France

A common task in food product development is to identify groups of consumers that are as homogenous as possible according to their responses (e.g., consumer perception of products, consumer preferences). In some cases, the segmentation based on consumer perception is not always consistent with the one on consumer preferences. When each consumer (i.e. each row in data set) is vector, the segmentation of conjoint measurements could be done with the help of clusterwise regression. One, however, faces with difficulties if each consumer is matrix, for example, rows are products, columns are response and exploratory variables. In this context, this study proposes a new approach for clustering consumers; first performing a PLS regression of liking on CATA data for each consumer, and then clustering consumers based on PLS regression coefficients.

Eight yoghurt samples, based on DOE, with identical composition, varying in textural properties, were used in the study. Consumers (n=101) tasted each sample, rated their liking and answered a Check-all-that-apply (CATA) question.

Many different tools can be applied for clustering based on the regression coefficients. Here we apply Fuzzy C-means (FCM). Using the fuzzy silhouette index, the FCM points out two clusters: cluster 1 (n=56), cluster 2 (n=45). Several analyses were applied to highlight the similarities and differences between clusters.

Correspondence analysis result displays that product separations are similar between clusters. The first component is characterised by viscosity whereas the second component by particle-size.

Considering overall liking, products that are high in viscosity are most liked. The main difference between clusters is the influence of flavour intensity on liking. While high in flavour intensity results in increase in liking for cluster 1, that is not for cluster 2.

Penalty-lift analyses were applied for further investigation. For cluster 1, flavour attributes (*Sweet*, *Oat.flavour*, *Vanilla*) are positive drivers, and their influence on liking is also larger than textural attributes. In contrast, for cluster 2, textural attributes (*Thick*, *Smooth, Dense*) are pointed as key drivers, followed by *Sweet*, *Vanilla*. Compared with clustering on data sets separately, the results seem more comparable with liking than CATA clustering.

The clustering approach proposed here provides a useful way to cluster consumers based on both liking and CATA data. More research is needed in products with different complexity in liking and sensory descriptions, to point out the appropriate criteria for cluster validation.

P6 Evaluation of beef consumption habits and preferences of Romanian consumers

Gliga Cristina, Tofană Maria, Stan Laura
University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Cluj-Napoca, Romania

On the Romanian market, there are a variety of beef species, such as: Black Angus, Wagyu or Kobe but few Romanian consumers are aware of it. Even fewer consumers know that the breed of Romanian Spotted is suitable for meat consumption, not just for milk production. This shortcoming is probably justified by the lack of consumer information.

The aim of this study was to evaluate the beef consumption habits taking into account all the beef breeds available on the market and the main cuts of beef. To achieve this aim, a survey with 18 questions was distributed online to consumers (n=299, from which 25% were removed from data interpretation because they were non-consumers of beef). Most of respondents were women (75%) and this may be due to the fact that in Romania women are more involved in meal preparation than men. The participants responded about the following topics: (1) frequency of beef consumption, (2) motivation of beef consumption, (3) purchasing channels, (4) the most preferred beef cut, (5) selection criteria of beef cut, (6) preference of beef breed, (7) preferred degree of doneness, (8) most frequently consumed gastronomic preparations with beef, (9) inspirational channels for gastronomic dishes with beef and (10) demographic data. The data were analyzed on Microsoft Office Excel, 2010.

The results indicated that the consumers prefer beef meat for its taste (45%) and due to perception of beef as being a healthy (44%). Most of consumers prefer to buy meat from butchery (47%). This preference reflects in hierarchy of the selection criteria of beef in buying decision: first place was manufacturer with 40%, compared to maturation degree (36%), packaging (23%) and type of species (27%). Most of the respondents (90%) consume beef less than once in a month. This low rate of beef intake might be due firstly to high price of beef meat compared to other types of meat available on the market.

In conclusion, more consumer education is needed, with focus especially on the beef consumption quality, recipes with beef and quality criteria of beef. Labeling of the products with information about the beef cuts and appropriate cooking methods for the specific cut, might be helpful to increase the beef intake.

KEYWORDS: beef cuts, matured meat, consumer habits, consumption patterns

P7 Consumer acceptance of a ready-to-eat meal during storage using a home-use test

Maria Laura Montero[1], Dolores Garrido[2], Juming Tang[3], R. Karina Gallardo[4,5], Carolyn Ross[1]
[1]School of Food Science, Washington State University, Pullman, USA. [2]School of Economic Sciences, Washington State University, Pullman, USA. [3]Department of Biological Systems Engineering, Washington State University, Pullman, USA. [4]School of Ecomic Sciences, Washington State University, Pullman, USA. [5]Puyallup Research and Extension Center, Puyallup, USA

Conducting sensory tests in a context reflecting more natural consumption has grown in interest, reflecting the concept of ecological validity. A home-use test (HUT) is one method that provides a measure of ecological validity as the product is consumed in-home under common daily use circumstances, usually for several days. One product that benefits from being evaluated in-home are ready-to-eat meals (RTE). RTE meals must be processed so they are microbially safe, and a novel method to accomplish this is microwave-assisted-pasteurization (MAPS). Therefore, this study determined consumer acceptance of MAPS-processed jambalaya and a control (cooked frozen jambalaya) through an on-line HUT over a 12-week storage period. Paralleling the HUT, an online auction determined consumers' willingness to pay.

Consumers (n=50) of RTE meals evaluated MAPS-processed jambalaya stored at 2°C and a control (cooked frozen jambalaya stored at -31°C) after 2, 8 and 12 weeks of storage. Consumer acceptance of different sensory attributes was measured for both jambalaya samples over the storage period. Using an online survey, the participants responded to questions using a hedonic scale and Just-about-Right (JAR) scale. Also, the main household could choose to share the meal with a partner, with the partner's responses (n=21) to the survey questions also collected. After sensory evaluation, the participants participated in an online auction to bid on the meal they had just sampled - this information was used to estimate their willingness to pay for each meal. Sensory data analysis was performed using ANOVA, with mean separation using Tukey's HSD and penalty analysis.

Results showed that the processing method (MAPS vs. control) did not significantly influence the measured sensory attributes. No significant sensory changes were observed in most of the tested attributes of the jambalaya samples (control or MAPS) due to storage time; only flavor liking decreased over time($p<0.05$). This same trend was observed when the partners' responses were included in the data analysis. Most of the tested texture-related attributes were around the JAR point, but consumers mainly penalized chicken texture when they considered it was too chewy. For the meals, the bid mean values ranged from $3.48-3.74 for the MAPS-processed jambalaya and from $3.33-3.56 for the control, similar to the price of commercially available jambalaya meals. This study suggests that using HUT for evaluating consumers' acceptance of MAPS-pasteurized jambalaya and including partner participation, is an effective way of testing acceptance of a ready-to-eat meal in a more realistic context.

P8 Sensory evaluation of cold pressed oils

Laura Stan, Anisoara-Ioana Popa
University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Cluj-Napoca, Romania

Cold pressed oils gain high popularity especially in the mind of health-conscious consumers due to nutritional value (especially hemp oil which is known for its high content in Omega 3 and Omega 6). Additionally, cold pressed oils are highly appreciated for their sensory qualities compared to refined oils. It was the purpose of this study to evaluate the Romanian consumers' acceptability of the following cold pressed oils: hemp oil (HO), rapeseed oil (RO), flaxseed oil (FO) and walnut oil (WO). The study was performed at USAMV Cluj-Napoca, Laboratory of Sensory Analysis of Foods. Voluntary consumers were selected based on their availability and oil consumption habits. Oil samplings were prepared for each participant in plastic transparent cups (20 ml oil) labeled with three-digit codes. Water and unsalted bread were provided for palate cleansing between samples. The sensory

characteristics evaluated on the 9 point hedonic scale were overall appreciation, aspect, color, viscosity, smell, taste, aroma and after taste. Participants were asked to record the intensity of their perception for some positive and negative sensory characteristics using a 5 level intensity scale (1 - not at all, 5 - intense). The data were analyzed on Microsoft Office Excel and Consumer Check. Most of the participants declared that they would mostly use these oils for salads or as food supplements. The general tendency of hedonic values decreased in the following order: WO > RO > HO > FO. The aftertaste varied greatly between samples and the strongest aftertaste as intensity and duration was evaluated for FO. More research focused on developing a sensory lexicon for cold pressed oils can provide efficient tools for quality control and consumer education.

P10 On the applicability of ANOVA models for CATA data

Michael Meyners[1], Anne Hasted[2]
[1]Procter & Gamble Service GmbH, Schwalbach, Germany. [2]Qi Statistics Ltd, West Malling, United Kingdom

Check-All-That-Apply (CATA) questions return binary data from every subject on each product and attribute. To compare multiple samples, Cochran's *Q* test is widely used to analyze such two-way data, with pairwise comparisons being conducted using McNemar's test, the special case of Cochran's *Q* for two samples. In many applications, this approach is viable, however, in some, it has limitations. We consider the following ones to be of most practical interest: (i) incomplete and/or imbalanced data, which does not allow direct use of Cochran's *Q* (except if done via a laborious randomization procedure); (ii) samples based on an experimental design, e.g., with regard to ingredients, for which further modeling is desired; (iii) testing more complex hypotheses than just overall and pairwise sample differences (e.g. interactions of gender and samples). If an analysis of variance (ANOVA) would be valid, all these issues would be resolved at once. CATA data is obviously not normally distributed, and logistic regression models might seem a more natural remedy. Yet, we have encountered convergence issue in logistic regression, particularly in accounting for the subject effect (which is frequently a substantial contributor to the variation in binary scores). Comparing analysis using ANOVA and Cochran's *Q*, we noted *F*-tests for balanced designs to provide p values very close to those from Cochran's Q test, and they have been shown earlier to be at least approximately valid for RATA data. Also, Cochran himself (1950) already noted the similarity between results from *F*- and *Q*-tests even for rather small sample sizes.

We explore whether ANOVA indeed offers a robust model for CATA data. To that end, we determined the randomization distribution for the respective *F*- and *t*-tests for various (balanced complete) CATA studies. The respective distributions overlap closely, suggesting that ANOVA-based tests (overall and for pairwise comparisons) are indeed valid. A limitation is found in case of very low elicitation rates across products, which leads to distinct discreteness of the randomization distribution for paired comparisons and hence the potential for slightly invalid results. Yet, even in those cases, *t*-tests seem to provide a reasonable approximation for most purposes. We will further discuss the properties for smaller number of subjects as well as showcase applications for imbalanced data and modeling in a study using an experimental design for the products. We conclude with guidance on when the use of ANOVA for CATA data is reasonable.