PermAnova testing and Poisson Log-Normal modeling unravel how two traditional cheeses are distinguished through sorting and verbalization tasks

P-M Grollemund[a], L. Lenoir[b], J. Benoit[b], C. Chassard[b], C. Bord[b]

[a]Laboratoire de Mathématiques Blaise Pascal (LMBP), Université Clermont Auvergne – IUT de Clermont-Ferrand

[b]Université Clermont Auvergne, INRAE, VetAgro Sup, UMRF, 63370 Lempdes, France

Corresponding author: Cécile Bord

cecile.bord@vetagro-sup.fr

VetAgro Sup, 89 avenue de l'Europe, 63370 Lempdes, France.

Highlights

- Three types of panel were selected to perform a sorting task on uncooked cheeses
- PermAnova testing and Poisson Log-Normal modeling identified key variables in a verbalization task
- Poisson Log-Normal model shows value for analyzing counts data
- Verbalization is best explained by 'knowledge' × 'panel' variables

Keywords: PDO cheeses, Free sorting, Verbalization task, PLN model, PermAnova

Abstract

The aim of this study was to explore how two categories of uncooked cheeses (Salers *versus* Cantal) can be perceived differently according to how panelists sort them (free sorting task) and how they verbally describe them (verbalization task). We focused on determining the mechanisms underpinning the ability to distinguish the two cheeses by considering additional panelist information. Three types of panels (an expert panel, an intermediate panel and a novice panel) with different levels of cheese expertise performed a sorting task and a verbalization task on 10 cheeses. Data from the sorting task was analyzed using the DISTATIS method and data from the verbalization task was analyzed by using a basic correspondence analysis paired with an original approach mobilizing PermAnova and a Poisson Log-Normal (PLN) model.

The results showed that the cheeses' configurations are similar except between the expert and intermediate panels. Overall, none of the panels clearly separated Cantal from Salers cheeses. In the verbalization task, different types of panel used different sets of terms to describe the categories. The expert panel preferentially used descriptive terms related to flavour whereas the intermediate and novice panels both tended to use quantitative (intensity) and hedonic terms. Knowledge of the product space under study was found to be the variable that best explains the terms used by each panel. PermAnova testing and PLN modeling emerged as novel approaches for identifying the key variables that explain the use of terms in the description task.

1. Introduction

France offers a huge number of all kinds of cheeses, but only 45 of them have obtained a Protected Designation of Origin (PDO) label to promote the quality and protection of authentic regional products. PDO products are distinguished by traditional craft and singular sensory qualities based on adherence to strict production standards and specifications. Among the typical French PDO cheeses, Cantal and Salers are uncooked cheeses produced in the Massif Central area (France). The two cheeses are similar but not the same in appearance, and many consumers struggle to distinguish between them. However, experts confidently distinguish the two cheeses based on particular gustatory differences (Bérard et al., 2016). Salers and Cantal cheeses are two very distinct cheeses that differ on several points, from production process (conditions of cheese-making, ripening time) to the dairy systems used (milk origin, seasonal production, and so on) (Callon et al., 2005; Agabriel et al., 2004) that converge to shape the distinct and distinctive sensory properties of Cantal and Salers cheeses (Cornu et al., 2009).

The sensory properties of various products can be measured using various descriptive sensory methods, such as sensory profiling (Meilgaard et al., 2015). Sensory profiling serves to describe specific qualities of a product and quantify them on an scale of intensity using a trained panel (*ISO 13299:2016*). However, there are faster alternative methods for obtaining a sensory map of products. Although popular in psychology, the free sorting method has largely developed in sensory analysis. It is a simple and relatively intuitive method in which participants freely group samples into subsets according to their similarities, and it can be used to assess a large set of both food and non-food products (Deegan et al., 2010; Merlo et al., 2022; Holliins et al., 1993; Giboreau et al., 2001). This method thus affords a perceptive representation of a set of samples.

Data from sorting tasks is generally analyzed using multidimensional scaling (MDS). There are many methods similar to MDS, such as the DISTATIS method that combines classical MDS (Abdi et al., 2007) and STATIS that accounts for the variability of panelists (individual data). DISTATIS yields two types of maps: one for the panelists and one for the products. These maps correspond to data projections in a low-dimensional space computed such as to lose the least amount of information. The proximity between two points in the maps reflects a partial similarity, which means that these maps can be analyzed using the same rules as standard metric MDS or Principal Component Analysis (PCA).

To give meaning to perceptive space, previous works suggest adding an optional step to a free sorting procedure by asking the panelists to verbalize, i.e. use terms to describe the groups formed (Cartier et al. 2006, Blancher et al. 2007). Faye et al. (2004) argued that the combination of categorization and description makes it possible to obtain explanations about the dimensions of the perceptual map. Even though this method looks simple to set up, several points of the methodology needed to be discussed, including the type of panel selected (trained/untrained panel) and the vocabulary used to describe the products. Several studies have compared trained and untrained panels on categorization tasks (Faye et al. 2004; Parr et al. 2010; Lelièvre et al. 2008). In the wine domain, these studies have shown that experts are able to categorize wines according to grape variety. Through repeated exposure to wines from different categories, experts develop a kind of prototypical or central tendency that represents a sensory signature of these categories (Honoré-Chedozeau et al., 2019). Other studies have shown that experts are better in terms of discrimination, memory and description. Despite this better discrimination and description by experts, MSD with naive panelists nevertheless leads to similar results regarding product positioning. However, some authors have reported that trained or expert panelists tend to be more efficient in their description, and other authors have shown that the vocabulary used by untrained or novices panelists can be ambiguous, redundant, and less specific than that the vocabulary verbalized by experts (Chollet & Valentin, 2001). Likewise, Lawless, Sheng & Knoops (1995) showed that expert panelists with a good knowledge of cheeses used a wide variety of meaningful descriptive terms compared to untrained panelists. Overall, its seems that knowledge and expertise play a role in how panelists describe products or utilize descriptor words.

On the other hand, qualitative analysis of terms is still difficult to interpret due to the different semantics used and the variety of terms selected. Verbalization generates a great many terms, and so before an MDS method can be applied, it is often necessary to pre-process the verbalization data, which can be done by categorizing the terms by family (Chollet & Valentin, 2000; Soufflet et al., 2004). As pointed out by Chollet & Valentin (2001), the description of products can mobilize several dimensions (hedonic, qualitative, quantitative), but the emergence of these dimensions depends on the coherent product space.

The dimensions of a product map can often be correlated with additional attributes to facilitate interpretation. The verbatim comments resulting from this categorization can then be analyzed by any of several methods, such as correspondence analysis which allows to project the cited terms on the map. However, other statistical methods could be used to better understand the use of word categories by experts/novices. In particular, the heterogeneity in verbalization can be reformulated as a hypothesis testing problem or a choice of model. In this context, permutational ANOVA (PermAnova) (Anderson, 2001, 2017) is a suitable tool. Moreover, using Bayesian Information Criterion(BIC) (Schwarz, 1978) to select a model from one of the various Poisson Log-Normal (PLN) regression models (Chiquet et al. 2021) is another suitable direction.

To our knowledge, no studies have used a statistical approach based on PermAnova testing or PLN modeling to investigate the effect of types of panel on the categorization of French PDO cheeses and how panels use terms to describe cheeses according to their expertise. In this context, the objectives of this study were to (i) identify the perceptual representation of cheeses by panelists according to panel expertise level, and (ii) determine the difference in verbalization (use of word categories) by different panels using PermAnova and PLN modeling to find out which variables (knowledge, cheese type and cheese consumption) best explain the use of the terms by the different panels.

2 Material and methods

2.1 Cheese samples

Two categories of cheeses were studied, i.e. Cantal (C) and Salers (S), which are specific PDO cheeses from the Auvergne region of France. For each category, cheeses were made with raw milk from either Salers (CSal/SSal) or other cow breeds (CR/SR). A total of ten cheeses were analyzed for this study (Table 1). All samples were stored in the same maturing cellar and were analyzed at 12 months of ripening. This process was implemented to precisely control the ripening conditions of all the cheeses. Whole cheeses (approximate weight of 40 kg) were selected directly in the maturing cellar by supply-chain professionals, and held in cold storage at 1°C for one week prior to analysis.

In order to be representative of the existing sensory diversity, three subsamples were selected for each of the stated categories of cheeses, except for CSal where a single subsample was analyzed.

2.2 Panel selection

Knowledge can be defined through two major components: familiarity and expertise. Iba & Hutchinson (1987) defined familiarity the number of product-related experiences the consumers have accumulated. More recently, Nacef et al. (2019) confirmed the different definitions of familiarity as related to product knowledge, experience with the product or the purchase, and consumption of the product. Expertise is defined as the ability to successfully perform product-related tasks. According to Chi, Feltovich & Glaser (1981), expertise is domain-specific and can be used to describe skills, knowledge or abilities in tasks, jobs, games or sports. In this case, ‘expert panel’ is not associated to sensory experts as defined in the ISO standards on sensory testing but to professional experience with cheese. Three sets of panelists were recruited:

- **an expert panel (n = 18)** including cheese-making professionals (farmer, cheese-maker, laboratory technician, cheese engineer or scientist) located around the cheese production area. In this case, experts (such as oenologists, brewers, cheese-makers, etc.) use their technical knowledge-sets.

- **an intermediate panel (n= 10)** including people working at Rungis market (a bastion of French gastronomic heritage and a well-known landmark in the gastronomic landscape), specifically people engaged in dairy sales, such as wholesalers or market operators.
- **a novice panel (n =44)** including cheese consumers living around the city of Clermont-Ferrand (familiar novices who regularly eat Cantal and/or Salers cheeses).

Participants were given questionnaire including items on socio-demographics (age, gender and professional activity) and tasting and consumption habits. To be selected for the consumer panel, consumers had to routinely consume Salers or Cantal cheeses and be over 18 years old. Informed consent was obtained from each participant. Table 2 describes the panels.

2.3 Experimental design

2.3.1 Free sorting and verbalization task

Different sessions were set up over three weeks, but each panelist only participated in a single 1.5-hour session. The sessions took place in different French cities according to panel type (expert panel: Aurillac and Lyon; intermediate panel: Rungis (Paris); novice consumer panel: Clermont-Ferrand). The sessions were organized in a room that was amenable to tasting (not a sensory laboratory) while adhering to the conditions required for sensory tests.

Panelists received a full set of cheese samples. A randomized plan gave the order for sample presentation. Three cubes of cheeses (2×2×2 cm) were cut and placed in a plastic cup. All cheeses were presented in three-digit numbers and served at room temperature (20°C±1°).

The panelists were asked to look at, smell, and taste each cheese sample to sort them into groups of cheeses based on sensory similarities. No criteria were provided to perform the sorting task. Number of groups and number of samples per group were free, as were the criteria for grouping the samples, which could vary from panelist to panelist. However, at least two groups had to be formed, and each cheese sample had to selected once into a group. After the sorting task, the panelists were asked to describe each group of cheeses using words (no sentences). Panelists were provided with mineral water and unsalted crackers to cleanse their palate between samples.

2.3.2 Knowledge questionnaire

A questionnaire was designed to assess the panelists' knowledge on cheese and validate their membership group. This questionnaire was split into two parts: (1) 13 questions related to their overall knowledge on cheese, (2) 12 specific questions about cheese-making process for Salers and Cantal cheeses. As the correct answer was known for each question, individual responses were coded 0 for a wrong response and 1 for a correct one. This questionnaire was filled out at the end of the tasting session, and was scored out of 20 points. A global knowledge index was obtained by computing a weighted mean of specific knowledge (weight: 0.7) and overall knowledge (weight: 0.3).

2.4 Data analysis

2.4.1 Analysis of the free sorting task

For each panelist, we built an individual distance matrix using the discrete metric from the results of the sorting task, for which the rows and the columns are cheese samples. Specifically, a value of 0 for the j-th element of the i-th line of the matrix indicates that the corresponding panelist pools the i-th and the j-th cheese samples together, whereas a value of 1 indicates that the samples were not put together. In this context, the dataset is a collection of individual distance matrixes on which an MDS method is usually used to derive sensory maps (Blancher et al., 2007). Here we used the DISTATIS method to analyze the individual distance matrixes obtained from the sorting data. DISTATIS method is an extension of classical multidimensional scaling taking into account individual sorting data (Abdi et al., 2007).

2.4.2 Analysis of the verbalization task

To facilitate the interpretation of the graphical displays obtained by MDS, the terms cited by panelists are superimposed by considering their correlations with the MDS axes. For each panelist, the terms are suggested to describe a given category of stimuli are assumed to apply to every single stimulus in that category. Very often, it becomes necessary to pre-process data from the verbalization task. Indeed, panelists generally generate a large number of separate terms, which can prove time-consuming to graph out. For instance, terms that are deemed to be synonymous or share a common meaning should be merged and identified as the same word. Sometimes, terms that are only cited by a few panelists generally get discarded.

Here, to properly encode each term used, we draw up a list of categories ready to integrate each term into a category prior to any data analysis. We chose to establish 9 categories to provide balanced and interpretable categories:

- Taste: all terms describing taste or intensity of perceived taste
- Texture: all terms describing texture (in-mouth or to-touch)
- Appearance: all terms describing appearance (core or rind)
- Odour intensity: all terms in relation to intensity of odour (quantitative terms)
- Flavour intensity: all terms in relation with intensity of flavour (quantitative terms)
- Odour-descriptive: all terms in relation with odour description (qualitative terms)
- Flavour-descriptive: all terms in relation with flavour description (qualitative terms)
- Liking: all terms describing preference, hedonic perception (liking or disliking)
- Typicity: all terms in relation to the perception of typicity (belonging to a given category of cheese, connected with the terroir or the ripening process)

In addition to a descriptive analysis performed via an MDS approach, a major objective of this study is to identify whether panelists' verbalizations depend on type of panel or on type of cheese to. Two approaches are investigated to test these potential links: a PermAnova test and a PLN model.

PermAnova is a non-parametric statistical test. Here, to evaluate the link between the type of panelists and their verbalizations, the null hypothesis corresponds to similar verbalization among panels in terms of mean and variance. Calculating the statistic involves computing the pairwise distances between all individual data. Given that verbalization data consists of vectors of count data, we choose to measure the proximity between two vectors using Bray-Curtis dissimilarity (Bray & Curtis, 1957). In this study, PermAnova is used to jointly test whether type of panel and type of cheeses resulted in a difference in verbalizations. Each p-value obtained is considered a measure of support for the null hypothesis, in line with recent scholarship on statistical significance (McShane et al., 2019).

As an alternative approach, the PLN method involves modeling the individual verbalization counts as a Poisson random variable for which the log rate has a prior individual-specific Gaussian distribution. Assuming an individual-specific prior distribution makes it possible to model potential heterogeneity in the intensity of verbalization across panelists. Moreover, adding regression terms to the prior Gaussian expectation leads to express the intensity of verbalization as a linear combination of dependent factors such as type of panel, panelist knowledge, and type of cheese. Here, we compared several PLN models with different regression terms to deduce which factors explain an important share of the variance in verbalization. Choice of model was based on BIC. The robustness of the proposed analysis was evaluated with a bootstrap procedure.

3. Results

3.1 Analysis of the free sorting task

3.1.1 Cheese configuration for each panel

Figure 1 shows the DISTATIS compromise maps of the 10 cheeses for each panel. Overall, we found slight differences in configuration between the types of panel, as confirmed by the values of the RV coefficient computed between the configurations of different panels. Configurations are globally similar between novice/expert (RV = 0.622, p = 0.007) and between novice/intermediate (RV = 0.525, p = 0.04) but are dissimilar between expert/intermediate (RV= 0.435; p = 0.153). None of the panels were able to distinguish between Cantal and Salers cheeses but all panels formed the same number of cheese categories, i.e. three groups, although there are differences between the groups identified.

For the novice panel, the first dimension (accounting for 32.57% of variance) opposed three Cantal cheeses (CR4, CR5, CR6) and two Salers cheeses (SSal7, SR10) against other cheeses, while the second dimension had two groups separating “SSal9/Csal3/SR12” from “SSal8/SR11”. The small confidence intervals reflect good agreement among panelists.

In the categorization performed by the intermediate panel, there is no clear opposition between cheeses and a greater overlap between confidence intervals, indicating weak agreement among panelists. The first dimension opposed Salers cheeses (SSal7, SSal8, SSal9, SR10) against a group composed of two Cantal cheese (CR5 and CR6) and one Salers cheese (SR11). the second dimension featured one group formed by three samples (SR12, CR4 and CSAl3).

In the categorization by the expert panel, the first dimension (accounting for 34.86% of variance) opposes four Salers cheeses (SR11, SR12, SSal8 and SSal9) against other cheeses composed of four Cantal cheeses (CR5, CR6, CR4 and CSal3) and two Salers cheeses (SR10 and SSal7). Like the novice panel, the small confidence intervals reflect good agreement among panelists.

3.1.2 DISTATIS robustness analysis

The numerical and graphical DISTATIS results point to two main interpretations. First, the intermediate panel appears to be significantly different to the expert panel in terms of cheese categorization. Second, the confidence ellipsis appeared to be larger for the intermediate panel compared to other panels (Figure 1). Note however that these interpretations may again be due to the small sample size of the intermediate panel. Below we address both these issues using bootstrapping and subsampling methods.

For the expert/intermediate RV coefficient, we randomly subsampled the expert panel and computed the ‘subsampled expert’/intermediate RV coefficient to obtain the same sample size for both panels and then evaluate how far sample size distinguished the intermediate and expert panels. We repeat the process 1000 times to obtain i) the proportion of p values greater than 0.01 and ii) the position of the original RV coefficient value (RV= 0.435) with respect to the empirical distribution of the RV coefficient under expert panel subsampling. There are 99.2% of p-value greater than 0.01, and the original RV coefficient corresponds to the 73.5%-quantile of the empirical RV distribution. Subsampling the expert panel has no effect on the difference between intermediate and expert panels: the two panels are significantly different in all cases.

To analyze the radius of the ellipsis for each panel, we compute the average rate between radius of ellipsis of the intermediate panel and the expert panel (resp. novice panel), i.e. 2.131 (resp. 1.465). As above, we randomly subsampled the expert and novice panels in order to obtain panels with an even sample size, then we repeated the subsampling method 1000 times to obtain the overall average rate between radius of ellipsis, i.e. 1.283 for the expert panel and 1.203 for the novice panel. Having a reduced panel sample size substantially decreased the radius of ellipsis, especially for the expert panel. Moreover, the original average rate corresponds to an extreme value of the empirical distribution of the average rate under subsampled panels, which also indicates that panel sample size drastically affects ellipsis radius. Nevertheless, even though the small panel sample size contributes to the ellipsis radius, the subsampled expert panel always led to a greater average ellipsis radius than the intermediate panel and in 99.9% of cases led to a greater average ellipsis radius than the novice panel. Furthermore, empirical distributions (for subsampled expert and novice panels) did not support the hypothesis that average ellipsis radius rate is even across panels, since it did not cover the value 1. Even though a smaller sample size increased the ellipsis radius, the intermediate panel still had larger ellipsis.

3.2 Analysis of the verbalization task

3.2.1 Main differences between panels

In this study, the panelists across all three panels used a total of 455 different words. Figure 2 represents the average frequency of the terms cited for each panel. Bigger balloon size reflects greater use of the term by the panel. Correspondence Analysis highlights across-panel differences in use of descriptive terms (Figure 3), which is confirmed by a Chi-squared test ($p < 2.2e-16$). The first principal component (89.9% of total inertia) contrasts the expert panel and other panels. The expert panel is strongly associated with descriptive terms (Odour-descriptive and Flavour-descriptive) whereas panelists from the intermediate panel preferentially refer to terms related to enjoyment of the eating experience (Liking). The second principal component (10.2% of total inertia) mainly to differentiated the intermediate *vs* novice panels. Although the novice panel is also associated with the Liking category, it is specifically related to other categories on sensation intensity (Odour intensity and Flavour intensity) and more particularly terms describing the visuals of the products (Appearance).

The on-average difference between panels can be evaluated by computing a contingency table then running a Chi-squared test and correspondence analysis. Intra-panel inertia can also be analyzed in order to measure the intensity of the split between panels.

3.2.2 PermAnova-based differences in verbalization

To compare panels on differences in product description, PermAnova is used as a non-parametric statistical test. Only pairwise distances between panelists are required to assess whether or not panels differed in terms of average and dispersion. Since each panelist’s data is a vector of word category counts, the distances between two panelists are computed according to Bray-Curtis dissimilarity. We find that the panels differed in verbalizations, whether in terms of average, dispersion or both (p = 0.001). PCA analysis is performed on the verbalizations table (Figure 4a) to identify whether average or dispersion (or both) differed across panels. According to the PCA results, only dispersion differs between panels.

The same process as above is also used to investigate differences in verbalization per type of cheese evaluated, i.e. Cantal *vs* Salers. The PermAnova test confirms a difference (p = 0.0051), which is also the case when jointly considering cheeses and panels (panel factor p-value = 0.0001 and cheese factor p-value = 0.0032). PCA analysis (Figure 4b) indicates that averaged verbalization does not differ between types of cheese but that there is slightly higher dispersion in verbalization around Salers.

For the type of panel and/or type of cheese factors, the PermAnova test finds significant differences that translate as a combination of differences in terms of average and especially dispersion.

3.2.3 Probabilistic modeling to assess panel differences

An alternative approach to assess verbalization differences between panels and or types of cheese is to formulate the question as a choice of model. We model vectors of counts, as abundance data, with a multivariate Poisson distribution for which the log-intensity is assumed to be Gaussian (Chiquet et al. (2021). The regression terms can be considered to relate heterogeneity in verbalization to independent variables, such as a panel variable that indicates which panel each vector data refers to. We test models that only included an intercept, or a panel regression term, or a product regression term, or panel and product regression terms, and so on for all the available independent variables.

The fitted models are compared based on BIC, and we use a bootstrap procedure to evaluate whether or not observed differences in BIC significantly diverged from to differences in BIC obtained from augmented models under additional pure random variable. Figure 5 shows that including panel factors in a PLN model induces a clear improvement in BIC, whereas the cheese factor does not appear relevant for explaining heterogeneity in verbalization, which means that the panelists generally fail to differentiate Salers from Cantal cheeses.

3.3 Knowledge and verbalization

Expert knowledge on cheese is mainly a key factor to drive verbalization heterogeneity among expert panel. Figure 5 shows that there is a hierarchical order of knowledge among the panels. Here, the three groups that we formed essentially on the basis of their experience and expertise with cheese were validated by their scores on the knowledge test. The expert panelists scored highest on overall knowledge and had similar scores to other panelists on both knowledge of cheese in general (Figure 5a) and more specific knowledge on Salers cheese (Figure 5b). Conversely, the novice panelists scored lowest on aggregate knowledge (9.5/20), with low scores for specific knowledge on Salers cheese and an intermediate score of 14.5 for knowledge on cheese in general, which is not significantly different to the intermediate panel.

According to the BIC (Figure 6), the PLN regression model that best explained the intensities of the categories verbalized was the model that factored in the knowledge and panel variables. The increase in BIC after adding a regression term related to the knowledge factor supports an influence of this knowledge on verbalization. Each type of knowledge has a similar BIC gain, which seems to indicate that either the types are highly correlated or their differences are not significant in terms of explaining panelists' verbalizations.

Moreover, the fact that the interaction of knowledge and panel variables leads to the best model of the diversity in verbalization across panelists means that:

- the panel variable alone is not sufficient, and that intra-panel diversity in knowledge has to be addressed as a key factor, and

- the knowledge variable alone is not sufficient, which means that two panelists who have roughly the same knowledge score but are in different panels might use different terms to describe cheese.

Table 3 illustrates an assessment of the average contribution of each variable for the best PLN regression model according to type of panel and panel knowledge. More specifically, it gives the average terms per panel, variable-by-variable, in the latent regression model predicting log intensity in the Poisson model of verbalization data. Positive (resp. negative) value indicate that the focal variable positively (resp. negatively) affects intensity of usage of that word-category. To facilitate interpretation, contributions of the ‘Salers knowledge’ variable are distributed across the different panels, as knowledge differs significantly from panel to panel ($p < 0.001$ for each pairwise panel comparison).

4. Discussion

4.1 Sorting task and terms used during the verbalization task

First, the categorization carried out on the 10 cheeses showed that even though some products were not grouped together, all between-panel configurations except expert/intermediate were fairly similar, but not identical. Overall, these results are in line with the literature on other products (Lawless & Glatter, 1990). For instance, Faye et al. (2013) found that configurations/sorting on wine glasses differed according to subject experience and knowledge.

The second objective of this study was to unravel the difference in use of the terms cited during a verbalization task. Several studies have shown that the combination of free sorting and verbal description leads to a perceptual map with explainable dimensions (Popper & Heymann, 1996). Several authors have shown that the dimensions of MDS configurations resulting from evaluation by untrained panelists can be interpreted using the vocabulary generated (Chollet & Valentin, 2000, Lim & Lawless, 2005, Lawless, 1989, Lawless et al., 1995). Giboreau et al

(2001) showed that perceptual maps on different fabric samples were relatively similar between trained and untrained panelists. However, other authors have shown that descriptions produced by untrained panelists are not always comparable to descriptions produced by expert subjects, as trained/expert panelists tend to be more efficient in their description. Lawless, Sheng & Knoops (1995) found that many of the descriptors used are significant but that experts use a wider variety of significant terms than untrained subjects. Furthermore, trained panelists use a more precise vocabulary as attributes generated by consumers are more ambiguous or redundant and less specific than attributes assigned by experts (Chollet & Valentin, 2001; Chollet & Valentin, 2006). In this study, we observe a similar pattern: after gathering the terms into word categories, we find that the expert panel has more words that are related to qualitative terms and therefore more specific, whereas the novice/intermediate panels used more terms related to pleasure (hedonic), product appearance, and intensity descriptors. Lawless et al. (1995) applied this same method on cheese samples and showed that the results were similar between trained and untrained panels, but the trained group had a greater number of significant attributes when regressed with the MDS space.

In the literature, several studies converge to agree that overall, experts present a more technical and specific vocabulary and they have the most extensive vocabulary than novices. These studies have mainly been conducted in the wine and beer sector (Chollet & Valentin, 2000; Gawel, 1997; Lelièvre et al., 2009).The superiority of the experts' vocabulary is due to the addition of technical and specific terms on top of the basic novice vocabulary, as reported in the extant literature. Indeed, several studies have shown the effect of expertise (knowledge) on the description of products. For instance, Lelièvre (2010) showed that beer professionals and beer connoisseurs perceived beer according to sensory, hedonic and technical properties whereas non-connoisseurs mainly structured their perceptions on the basis of sensory attributes. Likewise, Faye et al. (2013) showed that description and sorting on wine glasses were influenced by panelists’ knowledge and experience and their engagement with the products. Note, however, that they grouped their experts/non-connoisseurs post-hoc, whereas here we defined our expert/intermediate/novice panels beforehand. Even if frequency of consumption features in the definition of familiarity and plays a role in knowledge of a product, as mentioned by Nacef on Maroilles cheese, the utilization of terms here was not dependent on frequency of Salers consumption and this variable did not improve the PLN model.

4.2 Relevance of PermAnova and PLN modeling for the purpose of verbalization analysis

The novel statistical approach employed provided key insight into which variables drive the use of different terms for describing the products. We showed that different types of panels can produce similar perceptual maps but that the sorting and especially the word categories used depend on the panelists’ expertise and knowledge of the product. These results confirm those already illustrated in the literature, but the combination of PermAnova testing and PLN modeling offers a powerful new way to analyze data resulting from free sorting task and verbalization tasks.

PermAnova makes it possible to distinguish several levels of differences in verbalization. Graphical checking via PCA analysis can be used to determine whether verbalization data differ in average, in dispersion, or both. However, it would be instructive for future work to quantify the contribution of each component, i.e. average differences and dispersion differences, towards rejecting the null hypothesis.

PLN modeling brings a new way to study the variables that contribute to the use of word categories by panelists and thus better understand the mechanisms involved in perception in interaction with panelist knowledge as an experimental factor, as suggested by Faye et al. (2013). This approach makes it possible to capture which factors impact verbalization and, crucially, to determine the intensity and direction of the association. However, further developments can be envisaged for this PLN model. In particular, it may be useful to adopt a variable selection approach (regularized estimators) or to define a suitable prior in a hierarchical Bayesian model (spike-and-slab) in order to more robustly identify the relevant factors that explain differences in verbalization among panelists. In addition, the PLN model, which is used more in the field of ecology (Chiquet, Mariadassou & Robin, 2021) has received recent gained broader attention, and extensions have already been proposed. For instance, Zero-inflated PLN allows to model counts data with a large amount of zeros in the database. In the context of verbalization analysis, this approach could make it possible to analyze verbalization data without necessarily having to resort to creating word categories.

4.3 Limitations and directions for future work

Concerning the limits of our study, the size of the panels might be a weakness that could have an impact on the validity of the results. The panels used had different sizes depending on the type of panelists recruited or their training, and were relatively small, at between 10 and 20 panelists (Soufflet et al., 2004) when panelists are considered as professionals/experts. In contrast, the size is larger (ranges from 9 to 300) when panelists are untrained or naive consumers (Cadoret et al., 2009; Moussaoui & Varela, 2010; Varela & Ares, 2012, 2012). According to Faye et al. (2006), the number of consumers really necessary to perform the free sorting task may be small (here, 25 consumers still seems acceptable). However, we addressed the small sample size issue and we provided insights that support the results reported.

Regarding the verbatims, we did not study the loss of information according to number of consumers, but the number of terms increases with number of consumers and thus occurrences of the terms. The fact that we used representative consumer terms makes the explanation of the perceptive configuration robust, and the number of consumers recruited in this study enabled us to collect a large number of words.

5. Conclusion

The free sorting method followed by a verbalization task on 10 uncooked Cantal/Salers pressed cheeses revealed consistent configurations across different types of panels, except between the expert/intermediate panels. Overall, statistical analysis does not find a clear separation between Cantal and Salers cheeses, even in expert panel. Analysis of verbalization based on a combined PermAnova and PLN model approach highlighted that verbalization was differentiated according to type of panel. Furthermore, this approach showed that the 'cheese' variable did not appear to be a driving factor in panelist verbalizations. This indicates that panel verbalizations were not based on the nature of the cheese (Cantal or Salers) but rather on the 'knowledge' variable. This insight enabled us to more accurately model the verbalization data, and learn that the 'knowledge × panel' variables interaction better explained panelists' verbalizations or use of word categories. The PLN model method offers an informative way to analyze counts data, particularly from verbalization, and brings supplementary information to help better understand which variables best explain the choice of terms used by different panels.

Table

Table 1: Description of the cheese samples

Cheeses	Type of milk	Producer / Dairy	Cheese samples (code)
Cantal	Salers milk	Dairy 1	Csal3
Cantal	Other cow breed milk	Dairy 1	CR4
Cantal		Producer 1	CR5
Cantal		Producer 2	CR6
Salers	Salers milk	Producer 3	SSal7
Salers		Producer 4	SSal8
Salers		Producer 5	SSal9
Salers	Other cow breed milk	Producer 6	SR10
Salers		Producer 7	SR11
Salers		Producer 8	SR12

Table 2: Key characteristics of each panel

Cheeses	Modality	Expert (n =18)	Intermediate (n =10)	Novice (n =44)
Cheese_overall consumption	Several times/day	47.1	40.0	20.5
	1 time/day	29.4	10.0	36.4
	1 to 6 times/week	23.5	30.0	40.9
	2 to 3 times/month	0	10.0	2.3
	Never	0	10.0	0
Cantal_Consumption	Every day	5.9	0.0	4.5
	1 to 6 times/week	29.4	20.0	38.6
	1 to 3 times/week	47.1	40.0	47.7
	Never	17.6	40.0	9.1
Salers_Consumption	Every day	11.8	0.0	0.0
	1 to 6 times/week	47.1	30.0	15.9
	1 to 3 times/week	29.4	20.0	70.4
	Never	11.8	50.0	1.6
Gender	Male	72	80	39
	Female	28	20	61
Age	19-29	18	0	12
	30-45	29	50	23
	46-59	35	50	23
	> 60	18	0	42
School level	Secondary	17	20	45
	Undergraduate	56	40	33
	Masters	28	10	21
Marital Status	Single	17	20	18
	Married/Partner	83	80	59
	Separated	0	0	23

Table 3: Average contribution of each word category to the best PLN regression model according to type of panel and type of knowledge

Word_category	Novice_ panel	Intermediate_ panel	Expert_ panel	Novice _S_knowledge	Intermediate _S_knowledge	Expert_ S_knowledge
Liking	0.4467[1]	0.1457	-0.5554	0.2280	0.3412	0.4719
Appearance	-1.4833	-2.6582	-61.6768	0.7876	1.1783	1.6299
Flavour_descriptive	0.9489	-0.1606	3.0391	-0.9029	-1.3507	-1.8685
Flavour_intensity	0.7114	0.0519	-0.3581	0.0495	0.0741	0.1025
Typicity_concept	-0.7774	-1.2107	-0.1999	-0.2934	-0.4390	-0.6072
Texture	1.2508	0.8734	0.9832	0.0054	0.0081	0.0112
Taste	0.0350	-1.7041	0.8985	-0.3165	-0.4734	-0.6549
Odour_intensity	-0.4077	-1.4491	-0.5852	-0.5024	-0.7516	-1.0397
Odour_descriptive	-1.2857	-59.2474	-1.1911	-0.2896	-0.4332	-0.5993

S= Salers; [1]Contribution expressed in log intensity

Figures

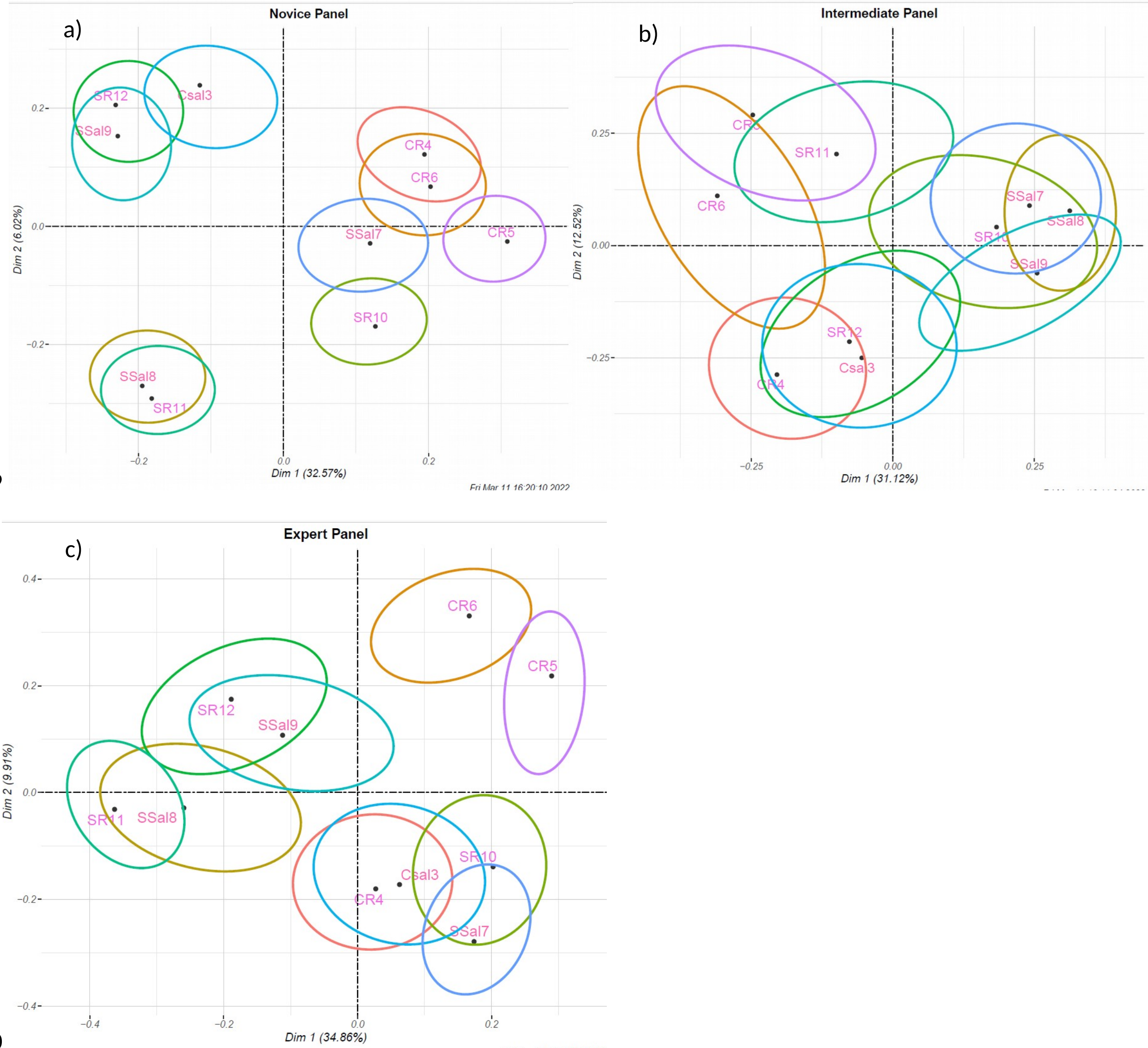


Figure 1: DISTATIS compromise maps of the ten cheeses with 95% confidence ellipses (Dim1-Dim2) for each panel: a) Novice panel (n= 44); b) Intermediate panel (n=10); c) Expert panel (n=18)

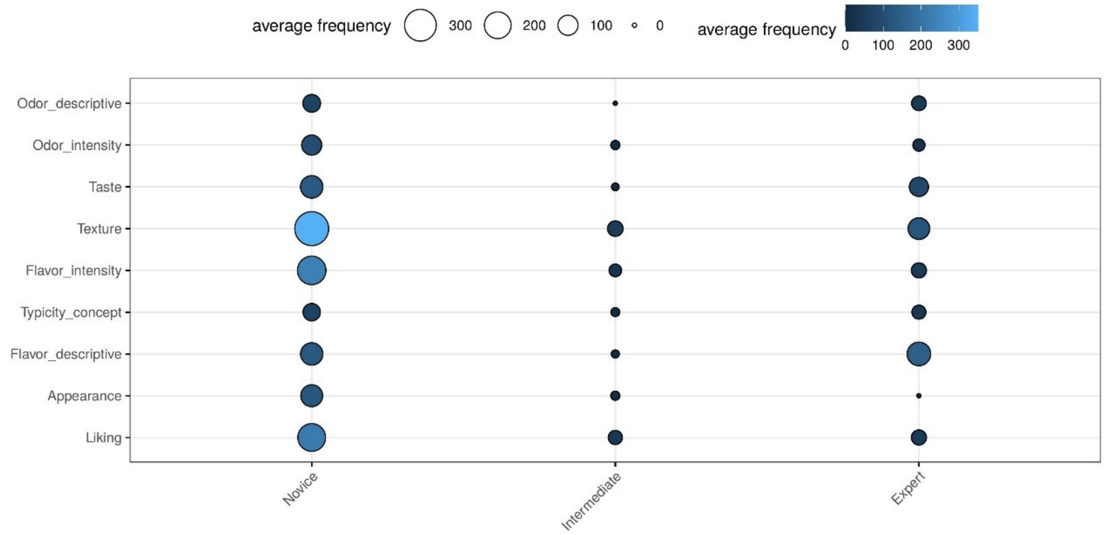


Figure 2: Balloon plot representing the average frequency of the terms cited for each panel

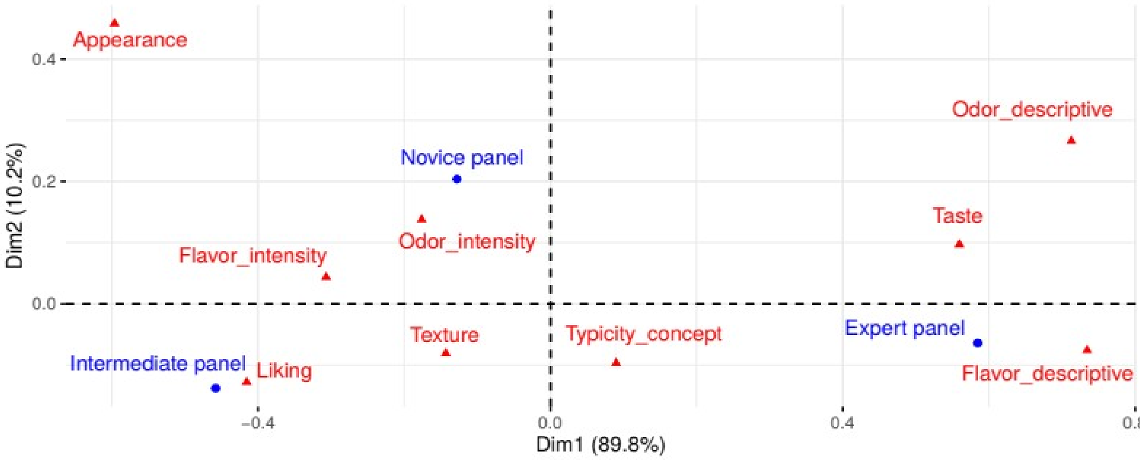


Figure 3: Plot from the correspondence Analysis between types of panel and cheese descriptor terms used

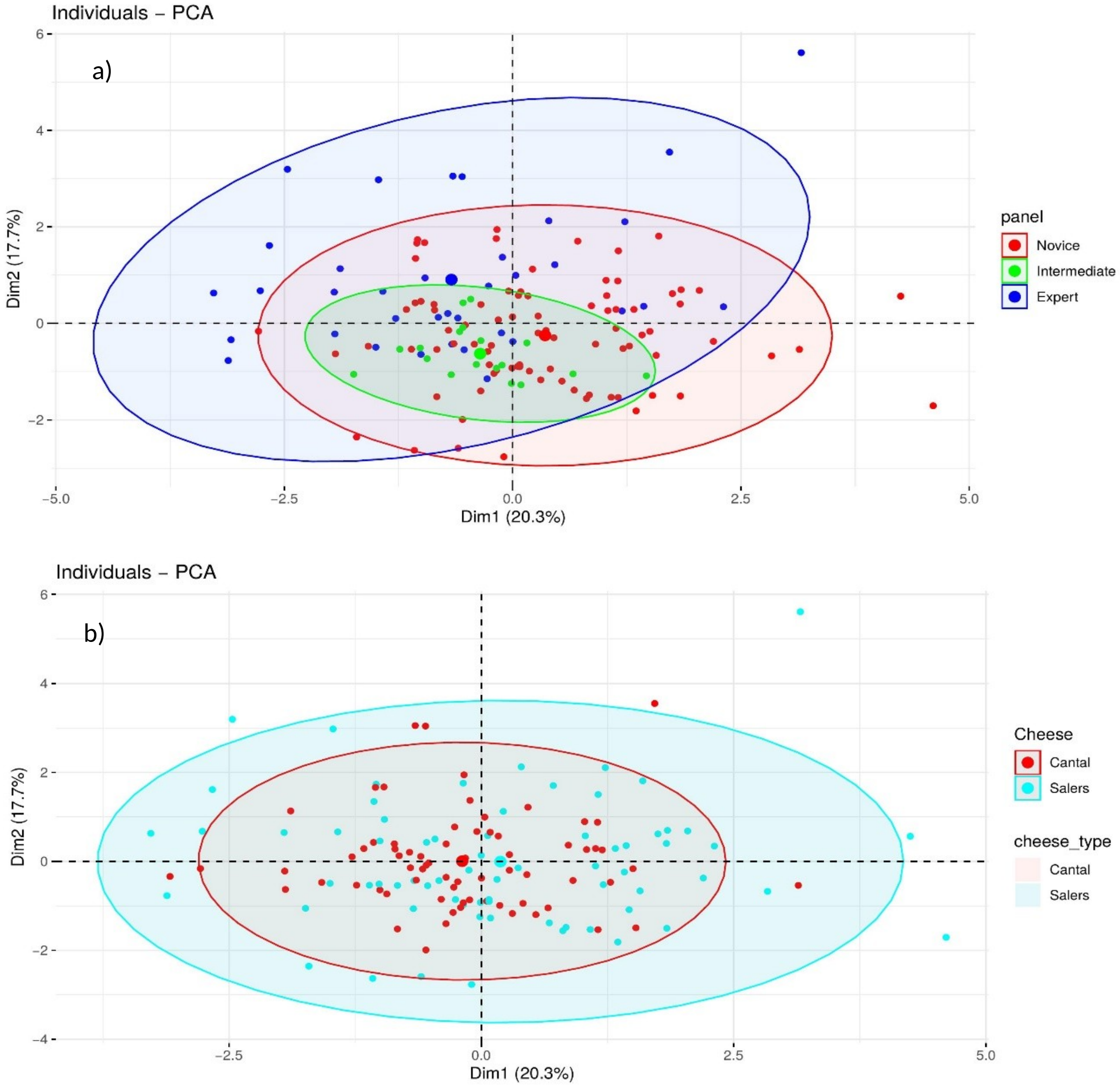


Figure 4: Plot from PCA performed on a) cheese categories and b) types of panel (Dim 1-2)

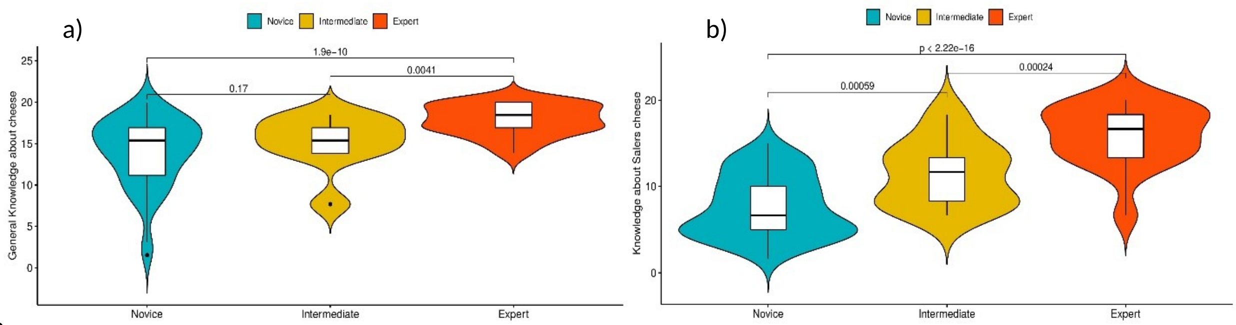
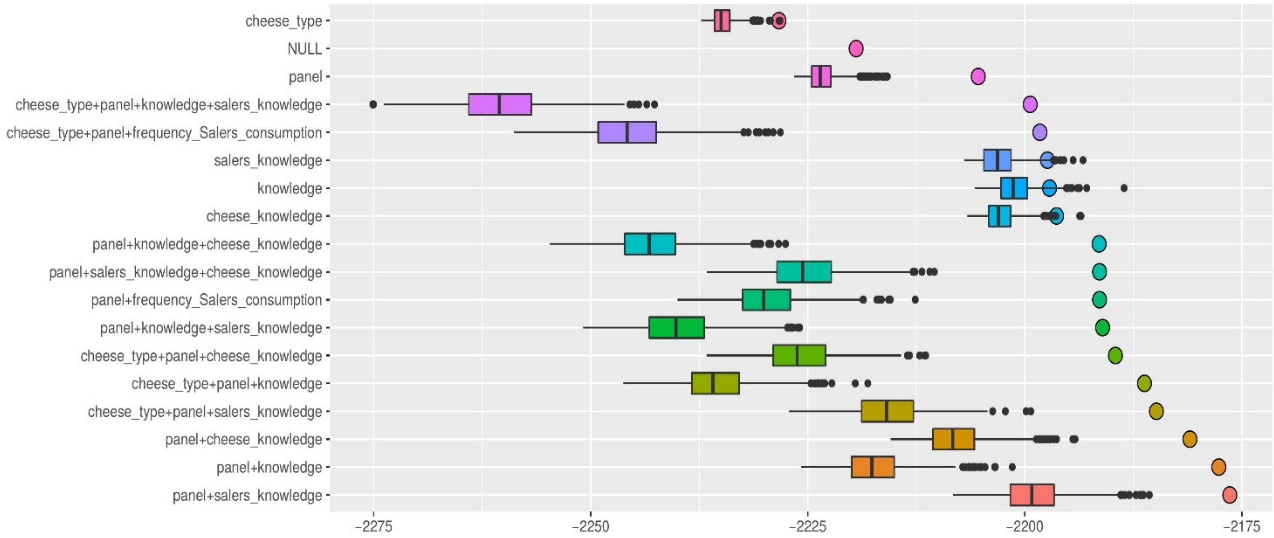


Figure 5: Scores on (a) cheese knowledge and (b) Salers knowledge for each panel



535

536 Figure 6: Plot graphing the BIC criterion according to the different PLN models analyzed

537

Sources of funding

This work was funded by FEADER and the Auvergne- Rhône-Alpes regional council under the rural development program for the project entitled “Innovation Fromagère pour Tradition Salers”.

CRediT authorship contribution statement

P-M Grollemund: Formal analysis, Writing – Original Draft, Writing – Review & Editing **C. Bord:** Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing. **L. Lenoir and J. Benoit:** Methodology. **C. Chassard:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank that many experts who took part in the study, the nonprofit ‘Tradition Salers’ for supplying the cheese samples, René Lavigne from INRAE-UMRF, and René Magneval (cheese-making experts) for their technical expertise. We also thank the consumers and participants in the sensory panels.

References

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference, 18*, 627-640. https://doi.org/10.1016/j.foodqual.2006.09.003

Agabriel, C., Martin, B., Sibra, C., Bonnefoy, J.-C., Montel, M.-C., Didienne, R., & Hulin, S. (2004). Effect of dairy production systems on the sensory characteristics of Cantal cheeses: A plant-scale study. *Animal Research*, *53*(3), 221–234. https://doi.org/10.1051/animres:2004013

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(3), 626–639. https://doi.org/10.1139/f01-004

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). In *Wiley StatsRef: Statistics Reference Online* (pp. 1–15). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat07841

Bérard, L., Casabianca, F., Montel, M.-C., Agabriel, C., & Bouche, R. (2016). Salers Protected Designation of Origin cheese, France. The diversity and paradox of local knowledge in geographical indications. *Culture & History Digital Journal*, *5*(1), e006. https://doi.org/10.3989/chdj.2016.006

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Cadoret, M., Lê, S., & Pagès, J. (2009). A Factorial Approach for Sorting Task data (FAST). *Food Quality and Preference*, *20*(6), 410–417. https://doi.org/10.1016/j.foodqual.2009.02.010

Callon, C., Berdagué, J.-L., Dufour, E., & Montel, M.-C. (2005). The effect of raw milk microbial flora on the sensory characteristics of salers-type cheeses. *Journal of Dairy Science*, *88*(11), 3840–3850. https://doi.org/10.1051/lait:198715

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices*. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chiquet, J., Mariadassou, M., & Robin, S. (2021). The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances. *Frontiers in Ecology and Evolution*, *9*. https://www.frontiersin.org/article/10.3389/fevo.2021.588292

Chollet, S., & Valentin, D. (2000). Le degré d'expertise a-t-il une influence sur la perception olfactive ? Quelques éléments de réponse dans le domaine du vin. *L'Année psychologique*, *100*(1), 11–36. https://doi.org/10.3406/psy.2000.28625

Cornu, A., Rabiau, N., Kondjoyan, N., Verdier-Metz, I., Pradel, P., Tournayre, P., Berdagué, J. L., & Martin, B. (2009). Odour-active compound profiles in Cantal-type cheese: Effect of cow diet, milk pasteurization and cheese ripening. *International Dairy Journal*, *19*(10), 588–594. https://doi.org/10.1016/j.idairyj.2009.04.008

Deegan, K. C., Koivisto, L., Näkkilä, J., Hyvönen, L., & Tuorila, H. (2010). Application of a sorting procedure to greenhouse-grown cucumbers and tomatoes. *LWT - Food Science and Technology*, *43*(3), 393–400. https://doi.org/10.1016/j.lwt.2009.08.014

Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: An alternative to descriptive mappings. *Food Quality and Preference*, *15*(7), 781–791. https://doi.org/10.1016/j.foodqual.2004.04.009

Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*, *17*(7), 604–614. https://doi.org/10.1016/j.foodqual.2006.05.006

Gawel, R. (1997). The Use of Language by Trained and Untrained Experienced Wine Tasters. *Journal of Sensory Studies*, *12*(4), 267–284. https://doi.org/10.1111/j.1745-459X.1997.tb00067.x

Giboreau, A., Navarro, S., Faye, P., & Dumortier, J. (2001). Sensory evaluation of automotive fabrics: The contribution of categorization tasks and non verbal information to set-up a descriptive method of tactile properties. *Food Quality and Preference*, *12*(5), 311–322. https://doi.org/10.1016/S0950-3293(01)00016-7

Holliins, M., Faldowski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics*, *54*(6), 697–705. https://doi.org/10.3758/BF03211795

Honoré-Chedozeau, C., Desmas, M., Ballester, J., Parr, W. V., & Chollet, S. (2019). Representation of wine and beer: Influence of expertise. *Current Opinion in Food Science*, *27*, 104–114. https://doi.org/10.1016/j.cofs.2019.07.002

ISO 13299:2016. Analyse sensorielle—Méthodologie—Directives générales pour l'établissement d'un profil sensoriel. (2016). AFNOR. https://www.iso.org/obp/ui/#iso:std:iso:13299:ed-2:v1:fr

Lawless, H. T., & Glatter, Sandy. (1990). Consistency of Multidimensional Scaling Models Derived from Odor Sorting. *Journal of Sensory Studies*, *5*(4), 217–230. https://doi.org/10.1111/j.1745-459X.1990.tb00492.x

Lawless, H. T., Sheng, N., & Knoops, S. S. C. P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, *6*(2), 91–98. https://doi.org/10.1016/0950-3293(95)98553-U lba, J. W., & Hutchinson, J. W. (1987). Dimensions of Consumer Expertise. *Journal of Consumer Research*, *13*(4), 411–454. http://www.jstor.org/stable/2489367Journal of Consumer Research

Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, *19*(8), 697–703. https://doi.org/10.1016/j.foodqual.2008.05.001

Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2009). Beer-Trained and Untrained Assessors Rely More on Vision than on Taste When They Categorize Beers. *Chemosensory Perception*, *2*(3), 143–153. https://doi.org/10.1007/s12078-009-9050-8

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2015). *Sensory Evaluation Techniques* (5th ed.). CRC Press.

Merlo, T. C., Saldaña, E., Patinho, I., Selani, M. M., & Contreras-Castillo, C. J. (2022). 10 - Free sorting task method to optimize the development of smoked bacon: A case study. In J. M. Lorenzo, M. Pateiro, E. Saldaña, & P. E. S. Munekata (Eds.), *Sensory Analysis for the Development of Meat Products* (pp. 173–179). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-822832-6.00010-2

Moussaoui, K. A., & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference*, *21*(8), 1088–1099. https://doi.org/10.1016/j.foodqual.2010.09.005

Nacef, M., Lelièvre-Desmas, M., Symoneaux, R., Jombart, L., Flahaut, C., & Chollet, S. (2019). Consumers' expectation and liking for cheese: Can familiarity effects resulting from regional differences be highlighted within a country? *Food Quality and Preference*, *72*, 188–197. https://doi.org/10.1016/j.foodqual.2018.10.004

Parr, W. V., Valentin, D., Green, J. A., & Dacremont, C. (2010). Evaluation of French and New Zealand Sauvignon wines by experienced French wine assessors. *Food Quality and Preference*, *21*(1), 56–64. https://doi.org/10.1016/j.foodqual.2009.08.002

Popper, R., & Heymann, H. (1996). Analyzing Differences Among Products and Panelists by Multidimensional Scaling. In *Data Handling in Science and Technology* (Vol. 16, pp. 159–184). Elsevier. https://doi.org/10.1016/S0922-3487(96)80030-X

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. JSTOR. http://www.jstor.org/stable/2958889

Soufflet, I., Calonnier, M., & Dacremont, C. (2004). A comparison between industrial experts' and novices' haptic perceptual organization: A tool to identify descriptors of the handle of fabrics. *Food Quality and Preference*, *15*(7), 689–699. https://doi.org/10.1016/j.foodqual.2004.03.005

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. https://doi.org/10.1016/j.foodres.2012.06.037