Published in final edited form as:

Food Qual Prefer. 2015 December 1; 46: 142–150. doi:10.1016/j.foodqual.2015.07.017.

Perception of chemesthetic stimuli in groups who differ by food involvement and culinary experience

Nadia Byrnes^{1,2,*}, Christopher R. Loss³, and John E. Hayes^{1,2,*}

¹Sensory Evaluation Center, College of Agricultural Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

²Department of Food Science, College of Agricultural Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

³Department of Culinary Science, The Culinary Institute of America, Hyde Park, NY

Abstract

In the English language, there is generally a limited lexicon when referring to the sensations elicited by chemesthetic stimuli like capsaicin, allyl isothiocyanate, and eugenol, the orally irritating compounds found in chiles, wasabi, and cloves, respectively. Elsewhere, experts and novices have been shown to use language differently, with experts using more precise language. Here, we compare perceptual maps and word usage across three cohorts: experts with formal culinary education, naïve individuals with high Food Involvement Scale (FIS) scores, and naïve individuals with low FIS scores. We hypothesized that increased experience with foods, whether through informal experiential learning or formal culinary education, would have a significant influence on the perceptual maps generated from a sorting task conducted with chemesthetic stimuli, as well as on language use in a descriptive follow-up task to this sorting task. The low- and highFIS non-expert cohorts generated significantly similar maps, though in other respects the highFIS cohort was an intermediate between the lowFIS and expert cohorts. The highFIS and expert cohorts generated more attributes but used language more idiosyncratically than the lowFIS group. Overall, the results from the expert group with formal culinary education differed from the two naïve cohorts both in the perceptual map generated using MDS as well as the mean number of attributes generated. Present data suggest that both formal education and informal experiential learning result in lexical development, but the level and type of learning can have a significant influence on language use and the approach to a sorting task.

Keywords

Perceptual mapping; chemesthetic; expert; culinary; spicy

*Corresponding author: Tel.: +1-814-863-7129; jeh40@psu.edu.

*current address: Viticulture & Enology, University of California Davis, Davis CA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. INTRODUCTION

Sorting is one approach within a family of methods commonly used to generate perceptual maps. Perceptual mapping techniques provide information about basic attributes and common characteristics that are relevant to the assessors, regardless of whether those assessors are experts or naïve participants. In a sorting task, assessors evaluate a group of stimuli and create groupings of the stimuli based on perceived similarities and dissimilarities. A related technique is napping, which takes its name from ‘nappe’ (tablecloth in French). In napping (also known as projective mapping), the participants place samples on a large sheet of paper so that similar samples are closer together and dissimilar samples are farther apart. Prior work suggests napping may provide better product differentiation when compared to sorting (Deneve & Cooper, 1998; Nestrud & Lawless, 2011) but there are significant drawbacks to the napping method. Specifically, the *ad libitum* retasting that is common in napping limits its utility with highly fatiguing samples, such as chemesthetic stimuli like capsaicin, zingerone, and menthol. Recently, we demonstrated that it is possible to conduct sorting with chemesthetic stimuli if the necessary precautions are taken (Byrnes, Nestrud & Hayes, 2015).

Generally, there can be substantial semantic confusion surrounding the sensations elicited by chemesthetic compounds (e.g. Bennett & Hayes, 2012). It is not uncommon to hear sensations that are easily distinguishable, such as those caused by chili peppers and horseradish, all colloquially referred to as being “spicy” or “hot” in spite of clear differences between them. Historically, one major advantage of using similarity-based judgments is that they avoid what Schiffman and colleagues termed “linguistic contamination” (Schiffman, Reynolds, Young, & Carroll, 1981). Here, we wished to explore the role of formal culinary education on the ability of assessors to differentiate between and describe the sensations elicited by a broad set of chemesthetic agents. We compared individuals with formal culinary education to naïve assessors as we reasoned that formal culinary education would enhance an individual’s personal lexicon regarding sensations from these various chemesthetic stimuli.

Previous work conflicts as to whether consumers are able to generate perceptual maps comparable to those generated by assessors trained via descriptive analysis techniques (here specifically called “trained panelists”) or individuals with specific expertise, both of which have been referred to as “experts” in the literature (Barcnas, Elortondo, & Albisu, 2004; Cartier et al., 2006; Chollet & Valentin, 2001; Faye et al., 2004; Faye et al., 2006; Gains & Thomson, 1990; Giacalone, Ribeiro, & Frøst, 2013; Guerrero, Gou, & Arnau, 1997; Kennedy & Heymann, 2009; Lawless & Glatter, 1990; Nestrud & Lawless, 2010; Pagès, 2005; Perrin et al., 2008; Risvik, Mcewan, & Rødbotten, 1997; Roberts & Vickers, 1994). The existing reports test the consensus of these configurations using a number of different methods, including sorting (Cartier et al., 2006; Lawless & Glatter, 1990), napping (Kennedy & Heymann 2009; Nestrud & Lawless, 2010), and free-choice profiling (Gains & Thomson, 1990; Guerrero et al., 1997), with a variety of stimuli including odorants (Lawless & Glatter, 1990), leather (Faye et al., 2006), beers (Giacalone et al., 2013), and cheddar cheeses (Roberts & Vickers, 1994). Critically, these reports also use a variety of different, and sometimes contradictory, definitions of the term “expert” (cf. Chollet, Lelièvre, Abdi, &

Valentin, 2011; Chollet & Valentin, 2001; Guerrero et al., 1997; Lawless & Glatter, 1990; Nestrud & Lawless, 2008; Pagès, 2005; Roberts & Vickers, 1994), perhaps accounting for the discrepancy in reported results.

In 1984, Lawless identified multiple types of different “experts” as “(1) trained panelists who use techniques such as the ‘Flavor Profile’ method, who have undergone a uniform and directed program of training, (2) persons who have such longstanding experience with a product that they are able to serve as ‘expert’ sensory evaluators, for example, in quality control work, and (3) persons who have made it their profession to develop new products based on sensory attributes, e.g., flavor chemists, perfumers, and the like” (Lawless, 1984). Even though trained panelists and experts are not equivalent (Perrin et al., 2008; Roberts & Vickers, 1994; Torri et al., 2013), the terms “expert” and “trained panelist” have been used imprecisely and somewhat interchangeably in prior literature, perhaps reflecting the varied usage described by Lawless. While these individuals are trained in different ways, and with different intent, experts and trained panelists do share key characteristics regarding their lexical and memory capacities. Perhaps due to these similarities, experts and trained panelists perform similarly on sorting tasks (Lawless & Glatter, 1990).

While acknowledging that domain specific experts and trained panelists are not identical or interchangeable, several common characteristics of experts, broadly defined, warrant comparison of their perceptual maps to those of untrained assessors. A few of the key differences identified in previous literature include differences in sensory acuity, language use or memory between experts and non-experts, or differential focus as a result of training. It has been proposed that experts may perform tasks differently than untrained assessors due to superior memory abilities, resulting in less of an impairment by a delay between samples and a better ability to cope with the memory load required in tests with repeated tasting of samples (Almeida, Cubero, & O’mahony, 1999; Chollet, Valentin, & Abdi, 2005; Nestrud & Lawless, 2010; Parr, White, & Heatherbell, 2004). Dissimilarities between experts and novices have also been attributed to differential use of language. While untrained assessors tend to use vague, less specific terms, experts and trained panelists tend to use language more precisely and efficiently (Chollet & Valentin, 2001; Clapperton & Piggott, 1979; Faye et al., 2004; Gains & Thomson, 1990; Guerrero et al., 1997; Lawless, 1984; Solomon, 1990). Importantly, assessment of the precision of word usage by experts and trained panelists refers to both the specificity and repeatability of the words. Solomon argues that the precise use of language allows for more subtle discrimination between samples (Solomon, 1990), a view supported by literature showing better discrimination in experts and trained panels compared to untrained assessors (Lawless, 1984; Risvik et al., 1997; Roberts & Vickers, 1994; Tang & Heymann, 2002; Torri et al., 2013), although it is also possible that experts may choose to become experts because of differences in innate ability (e.g. (Hayes & Pickering, 2011)).

While trained assessors may be better at identifying small differences between samples, training or expertise may also shift the way assessors attend to the task (Delarue & Sieffermann, 2004; Roberts & Vickers, 1994; Torri et al., 2013). For example, Roberts and Vickers (Roberts & Vickers, 1994) observed that trained dairy judges focus on defects in cheeses, rating primarily negative qualities, as compared to assessors not trained in dairy

judging, who rated both positive and negative attributes. Likewise, Torri and colleagues (2013) showed that wine experts tended to generate napping configurations that were equivalent to quality assessments while consumers tended to sort based on hedonic criteria. Overall, it remains unclear whether data from experts are more reproducible or more idiosyncratic than consumers as past reports conflict (Barcenas et al., 2004; Nestrud & Lawless, 2008; Torri et al., 2013); we wished to explore this further here.

Bell and Marshall (2003) conceived of the Food Involvement Scale (FIS) as a general measure of overall involvement with food, where food involvement is defined as the level of importance of food in someone's life. Their scale measures involvement with food across five different stages: acquiring, preparing, cooking, eating, and disposal. Factor analysis indicates individual scale items load onto two subscales: preparation and eating (FIS-PE) and set up and disposal (FIS-SD). While an individual's moods or cravings may change throughout the day, making them more or less likely to want to prepare a meal versus dine out, Food Involvement is more similar to a personality trait in that it does not vary from moment to moment (Marshall & Bell, 2004). Marshall and Bell reported individuals with higher FIS scores show finer discrimination between food items, both in intensity and hedonic ratings (Bell & Marshall, 2003), although pilot work from our lab suggests this may not be a robust effect (Byrnes & Hayes, 2013). Regardless of differences in sensory acuity that may potentially exist across groups with different expertise (see Hayes & Pickering, 2011), we anticipate that higher interest in and experience with food, as measured with the FIS, will lead to greater learning about food. Thus, we also anticipate that highFIS individuals will fall between the lowFIS and expert cohorts regarding lexical development. To test the effect of formal culinary education, we recruited a group of individuals from a culinary school as a cohort of experts. We hypothesized that the expert cohort would perform a free sorting task with chemesthetic stimuli similarly to the non-expert cohorts in terms of the perceptual maps they created, but that they would outperform the non-experts in the descriptive portion of the task, where they provided verbal labels to describe the groups formed during the free sorting task.

2. MATERIALS AND METHODS

2.1. Overview

This study was performed in three separate cohorts of individuals. All conditions, stimuli, and instructions were the same in each cohort. All data were collected with the approval of the Penn State Institutional Review Board; all participants provided informed consent.

2.2. Participants

Participants were recruited from the Penn State campus and the surrounding area in State College, Pennsylvania as well as through the Culinary Institute of America in Hyde Park, New York. To be eligible, individuals needed to be non-smoking, fluent English speakers between 18 and 55 years old, with no known food allergies or defect of taste or smell. Additional exclusion criteria included being pregnant or nursing, having known difficulties swallowing, or a history of thyroid irregularities; individuals with any chronic pain condition requiring prescription analgesics were also excluded (e.g. Green & Hayes, 2004). As

perceptual maps from sorting stabilize with around 25–30 participants (Faye et al., 2006; Lawless & Horne, 2000), we tested at least 30 participants in each cohort. To qualify for the expert cohort, participants must have been culinary students beyond their second year of instruction, or instructors who had already attained a degree from a culinary institution. It is possible a few of our ‘naïve’ participants had also received some sort of formal culinary education previously, but we did not formally assess this in our data; nonetheless, we would expect the number is quite low, given the recruitment pool.

2.3. Stimuli

Samples were prepared in ethanol (95%, USP, Koptec, King of Prussia, PA), with the exception of citric acid and quinine, which were prepared in reverse osmosis (RO) water. All samples were Food Grade (FG), Food Chemical Codex (FCC), Kosher, or U.S. Pharmacopeia (USP). Eugenol (12.2mM), menthol (38.4mM), allyl isothiocyanate (0.36M), zingerone (vanillylacetone; 59.7mM), quinine (4.1mM), cinnamaldehyde (0.12M), and carvacrol (0.27M) were obtained from SAFC (St. Louis, MO), citric acid (112mM) from J. T. Baker (Phillipsburg, NJ), capsaicin (100uM) from Sigma, eucalyptol (0.65M) from International Flavors and Fragrances (Union Beach, NJ), and huajiao extract (red fraction; 5% w/w) was a gift from Dr. Christopher Simons (formerly with Givaudan in Cincinnati, OH). The stimuli concentrations used in this experiment were adapted from previous literature on oral delivery of chemesthetic compounds, and refined via iterative pilot testing with ~12 participants drawn from our research team. Using paper ballots and the tasting protocol described previously (Byrnes, Nestrud, & Hayes 2015), the final concentrations were selected as they elicited sensation intensities near “moderate” on a general Labeled Magnitude Scale (gLMS). Across the group, these concentrations were high enough to elicit a distinct sensation but low enough that the sensation fully dissipated in three minutes, although no specific attempts were made to account for individual differences.

In order to maximize the number of chemesthetic stimuli that were assessed while ensuring that stimuli covered a wide range of sensations, we included two tastants in the stimulus set. Quinine does not elicit burning or stinging (e.g. Green & Hayes, 2003); it was included because some individuals report bitterness from capsaicin and zingerone (Green & Hayes, 2003, 2004). Citric acid was also included in the stimulus set as a sour exemplar; however, we should note that citric acid can evoke irritancy when applied to the anterior tongue via 0.5 inch diameter filter paper disks at concentrations similar to those used here (Gilmore & Green, 1993). Given the greater area of stimulation here, it is unclear how much of the overall citric acid percept was sourness as opposed to irritancy.

2.4. Procedure

Stimuli were prepared as stock concentrations and kept up to three weeks. Cotton swabs were saturated in stock solution and dried, cotton end up, with the wooden shaft pressed into blocks of florist’s foam. Solutions in ethanol were dried for three hours and solutions in water were allowed to dry for 10 hours. Swabs were tagged with three-digit blinding codes and stored in plastic zip-top bags for up to one week.

Samples were presented in glass culture tubes, with two swabs in a tube for each stimulus. Participants used a dispensing pipette to pump 10 ml of mouth temperature (35C) RO water into a new medicine cup. They then placed the swab in the water until fully saturated: no minimum or maximum time was enforced for this step, but this process typically took 3 – 5 seconds. Participants then rolled the swab across their tongue three times, making sure to cross the center line, then rubbed the swab against the roof of their mouth three times, breathed in through their mouth three times, allowing air to pass over the tongue, and finally pressed the tip of their tongue to the roof of their mouth three times. Prior to rinsing with mouth temperature RO water, participants placed the sample into the group they thought was appropriate or they formed a new group. A three-minute minimum interstimulus interval was enforced, and participants rinsed *ad libitum* (at least twice) and waited until no lingering sensation was perceived before moving on to the next sample. Retasting with a fresh swab was allowed, but participants were required to follow the complete protocol as if they were tasting a new sample (i.e., pump new warm water, rewet new cotton swab to saturation, swab tongue 3 times, etc.). There was no limit placed on number of retastings or on time spent on a single stimulus. All samples and rinse water were expectorated.

As placeholders during sorting, participants used identically colored poker chips that had been labeled with three-digit codes corresponding to the blinding codes on the swabs. Participants were instructed to form groups of the samples based on perceived similarities and dissimilarities: however the criteria on which these grouping were formed were determined by the participants without additional instruction from the experimenter. Study participants were told at the beginning of the session that they did not need to name the groups that they formed. After they had tasted all of the samples and decided on a final configuration, participants input their groupings into a web-based card-sorting program, Websort, (UXPunk, Chicago, IL, USA; subsequently purchased by Optimal Workshop, Wellington, NZ and renamed OptimalSort; <http://www.optimalworkshop.com/optimalsort.htm>) and they were asked to provide a description of each group.

In addition to the stimuli, participants were given a notepad and pen to keep notes, a sheet with the sampling directions outlined, and a list of possible descriptors. Participants were reminded that this list was not a comprehensive list but could serve as a starting point if they wished to reference it. The list of words included anesthetizing, astringent, biting, bitter, burning, buzzing, cooling, drying, hot, irritating, itching, metallic, numbing, pricking, puckering, salty, sharp, sour, spicy, stinging, sweet, swelling, tickling, tingling, umami/savory, and warming. No definitions were provided. These words were chosen as a compilation of words previously applied to chemesthetic agents (Albin & Simons, 2010; Bennett & Hayes, 2012; Cliff & Heymann, 1992) with the addition of the five prototypical tastes to the list. The list was presented in alphabetical order for all participants.

The total session took participants approximately an hour, including consent, sorting, naming, and a questionnaire at the end, which included the FIS.

2.5. Data Analysis

Multidimensional scaling (MDS) was performed using The R Statistics Package (R Foundation for Statistical Computing). In R, we used the smacof library for MDS, the *agnes*

function in the cluster library for cluster analysis, and the FactoMineR (Husson, Lê, & Cadoret, 2014) library to calculate normalized RV coefficients to compare the perceptual maps (see below). Examples of the R code used are provided in supplemental materials.

Data from the free sorting task was converted into a dissimilarity matrix and submitted to MDS. To determine the appropriate number of dimensions for the perceptual mapping solution, we used a Scree plot with Kruskal's stress values as a function of the number of dimensions in the MDS solution. The appropriate number of dimensions was chosen as the point when an increase in dimensionality did not meaningfully decrease the stress of the solution or did not aid in the interpretation of the configuration. Generally, a Kruskal's stress level below 0.1 is considered an acceptable model fit (Krzanowski & Marriott, 1994). A blind duplicate was not included within the sample set to reduce the potential participant fatigue; however, for the purpose of assessing reliability of the perceptual maps, we consider zingerone and capsaicin as quasi-replicates.

The RV coefficient (Robert & Escoufier, 1976), a multivariate generalization of Pearson's R^2 , is commonly used as a measure of similarity between the multivariate configurations from multiple cohorts. Here, we used the normalized RV (NRV) coefficient, as the number of stimuli in a group and dimensions in the perceptual map can influence the RV coefficient (Nestrud & Lawless, 2008). The NRV is interpreted similarly to a z-score, with a large score (>2) indicating significant similarity between the maps. The *coeffRV* function in FactoMineR also computes a p-value that tests for significant similarity when comparing the perceptual maps.

Multiple regression was used to associate the descriptors that participants generated in the last step of the experimental protocol with the stimuli coordinates from the perceptual map. This allows us to visualize which attributes were significantly associated with what stimuli as attribute vectors in the perceptual maps, similar to (Schiffman et al., 1981). The six most frequently used attributes for each stimulus were regressed onto the MDS coordinates of the samples to determine which attributes were significant across each cohort. Descriptors with p-values less than 0.1 were considered significant in the regression. The top six attributes for each stimulus (by cohort) are given in the supplemental materials.

3. RESULTS

3.1. Panelist demographics

Non-expert participants were split into high and low Food Involvement Scale (FIS) groups via a median split. Figure 1 displays a histogram of each of the cohorts' total FIS scores. Scores on the FIS for the lowFIS cohort ranged from 44 – 66 (possible range 12 – 84), with the mean score equal to 57.3 (± 1.2 SE). For the highFIS group, scores ranged from 67 to 81 with the mean score 72.3 (± 0.8 SE). The expert cohort's FIS scores ranged from 58 to 84 with mean 72.8 (± 1.2 SE). There was a significant effect of cohort on mean FIS scores ($F_{2,79} = 59.8$, $p < 0.0001$). The expert and highFIS cohorts showed significantly higher scores on the FIS scale than the lowFIS group (both p 's < 0.0001); however, there was no significant difference in the FIS scores between the highFIS and expert cohorts ($p = 0.956$).

This effect was present in both of the FIS subscales, Setup and Dining (FIS-SD; $F_{2,79} = 11.91$, $p < 0.0001$) and Preparation and Eating (FIS-PE; $F_{2,79} = 44.33$, $p < 0.0001$).

Mean age of each of the cohorts was roughly 29 years old (lowFIS: 27.8 $\pm$ 1.7 years old, highFIS: 28.0 $\pm$ 5.7 years old, experts: 29.9 $\pm$ 2.2 years old). There was no significant difference in the mean age of the cohorts ($F_{2,79} = 0.41$, $p = 0.662$). The lowFIS and expert cohorts were roughly 50% male (53.8% and 54.8%, respectively), while the highFIS cohort was 32% male.

3.2. LowFIS cohort

Figure 2 shows the two-dimensional MDS configuration for assessors with low scores on the Food Involvement Scale (FIS). As determined using a Scree plot, a two-dimensional solution was most appropriate for this data (stress = 0.008). As expected, zingerone and capsaicin fell very close to each other, suggesting the map is reliable.

The lowFIS cohort generated 35 unique attributes in total, of which 18 were submitted to regression, and eight were significant in regression analysis. The significant attributes were savory, herbaceous, puckering/sour, anesthetizing/numbing, cooling, warming, spiced, and spicy. There were two roughly orthogonal axes in Figure 2. The first axis opposes attributes cooling and anesthetizing/numbing with the attribute spicy. Along the second axis, the attribute puckering/sour opposes the attributes warming and spiced. There is a third unipolar axis that falls between the attribute vectors for puckering/sour and spicy. This dimension is made up of the attributes savory and herbaceous.

3.3. HighFIS cohort

The two-dimensional MDS solution determined to be the best model for the data generated by naïve assessors with highFIS scores is shown in Figure 3. Stress for this two-dimensional model was 0.008. Again, zingerone and capsaicin fell near each other, suggesting the map was reliable.

Although only two attributes were significant in regression, the highFIS group generated 56 unique attributes, of which 20 were submitted to regression. The two significant attributes lay on one opposing axis with the attribute astringent/drying on one end and the attribute herbaceous on the other end.

3.4. Expert cohort

The perceptual map for the expert cohort ($n = 32$) is shown in Figure 4 (stress = 0.014), and the quasi-replicates zingerone and capsaicin fell near to each other. The experts generated 54 unique attributes during the sorting task. Of these, 19 were submitted to regression, and only two attributes were significant in regression. The two significant attributes, cooling and anesthetizing/numbing, make up a single unipolar axis.

3.5. Comparing perceptual maps between cohorts

Normalized RV coefficients (NRVs) were calculated between the cohorts to provide a statistical measure of the similarity of the perceptual maps. The perceptual map generated

from the expert cohort's data was significantly different from both the low and highFIS groups (expert versus lowFIS: $NRV = 1.47$, $p = 0.08$, expert versus highFIS = 1.12 , $p = 0.13$). The highFIS and lowFIS cohorts' perceptual maps were significantly similar to each other ($NRV = 2.15$, $p = 0.03$).

To explore if there were differences in the way that assessors in the different cohorts conducted the task, we examined mean number of groups formed, mean number of attributes generated, and amount of overlap in the attributes that were used. Four levels of overlap were determined ranging from high (3) to none (0). An assessor who used the same term to describe more than two groups, or used three words twice or more was determined to have high overlap. Medium overlap was when the same terms were used in two groups or two words were used two or more times, low overlap was when assessors reused one word, and no overlap was when there were no descriptors that were reused. No significant difference was observed in the number of groups formed ($F_{2,78} = 1.732$, $p = 0.184$) or the amount of overlap ($F_{2,78} = 0.883$, $p = 0.418$). As summarized in Table 1, there was a significant difference in the mean number of attributes generated by assessors in each cohort ($F_{2,78} = 10.10$, $p = 0.0001$). Tukey's HSD indicated the expert cohort generated significantly more attributes than both the highFIS ($p = 0.013$) and lowFIS ($p = 0.0001$) cohorts. No significant difference was observed in the mean number of attributes generated between the lowFIS and highFIS cohorts ($p = 0.358$).

The lowFIS cohort generated 35 unique attributes during the descriptive portion of the sorting task as compared to 56 unique descriptors for the highFIS cohort and 54 unique descriptors for the expert cohort. The distribution of types of sensations across cohorts is shown in Table 2.

4. DISCUSSION

Here, we find evidence that the perceptual maps generated by experts differ from those generated by non-experts, in agreement with some prior reports. We expected that formal culinary education would enhance the food-related lexicon of the expert cohort when compared to the cohorts of naïve assessors, leading to significantly better performance on the descriptive portion of the task. Somewhat surprisingly, the highFIS cohort behaved similarly to both the lowFIS and the expert cohort in the descriptive portion of the task. Collectively, the results from the sorting and descriptive portions of this task suggest that formal culinary education did not differentiate assessors as distinctly as originally hypothesized. We observed that informal experience gained through increased involvement with foods does not alter the approach to the sorting portion of the task but that this experience does appear to have a large influence on an assessor's approach to completing the descriptive task.

While the perceptual maps may appear very different between the three cohorts on visual inspection, only the expert cohort's map is significantly different from the lowFIS and highFIS cohorts' maps when compared statistically using an NRV coefficient. Currently, there are conflicting reports regarding the ability of untrained assessors to generate maps that are similar to those by trained or expert assessors using perceptual mapping techniques.

Some prior work proposes untrained assessors cannot generate maps that are comparable to maps generated by experts (Barcenas et al., 2004; Nestrud & Lawless, 2010; Pagès, 2005; Perrin et al., 2008; Risvik et al., 1997) while other findings suggest untrained assessors are able to generate product maps comparable to those generated by trained or expert assessors (Chollet et al., 2011; Faye et al., 2004; Faye et al., 2006; Lawless & Glatter, 1990; Tang & Heymann, 2002). It is possible that these contradictory results arise from differences in the methodology performed (e.g., napping, sorting, or free choice profiling), type of training or degree of expertise of the trained/expert cohorts, and the type of stimuli used and size of perceptual differences between the stimuli within the sample set. These possibilities are discussed in more detail below. The incongruence between the maps of the expert and naïve assessors suggests that the expert cohort may be either a) attending to the sorting task differently or b) actually perceiving the stimuli in a different way than the naïve assessors. Indeed, during data collection, we informally noted that a number of the assessors in the expert cohort seemed to treat the task as an identification task, asking after the session if they had correctly identified the sensations' culinary sources; this behavior was never observed with the naïve assessors.

It was expected that the lowFIS group might have more difficulty discriminating between samples for two reasons, the first having to do with FIS scores and the second to do with the lexicon of these assessors. Previous work with the FIS has suggested that individuals with higher scores on the FIS have higher sensory acuity than those individuals with lower scores (Bell & Marshall, 2003). Given these findings, it might be expected that lowFIS individuals may have more difficulties discerning the perceptual differences between the chemesthetic sensations elicited by the stimuli in the sample set, irrespective of the descriptors they provided. Further, even if lowFIS assessors were able to pick apart these differences perceptually, we expected they would have a smaller food-related lexicon with which to describe the samples linguistically, which would create difficulties during the descriptive portion of the task. Contrary to our hypothesis, the lowFIS cohort did not appear to have more difficulty completing the sorting task as a group (stress = 0.008) than the highFIS cohort (stress = 0.008).

Generally, the significant attributes on each of the MDS plots associated with the expected stimuli. For example, on the lowFIS cohort's map, cooling and anesthetizing/numbing point towards menthol and eucalyptol, spicy points towards capsaicin and zingerone, and puckering/sour points towards citric acid. In addition to looking at the significant attributes, examining the number and type of attributes generated by each cohort provides potentially interesting information. As a group, the lowFIS cohort generated relatively few unique attributes, only 35, compared to the 56 and 54 unique attributes generated by the highFIS and expert cohorts, respectively. Although there were large differences in the number of unique attributes generated by the cohorts, roughly 20 attributes were submitted to regression for each cohort (see supplemental materials). From the regression analysis, eight attributes were significant in the lowFIS group while only two were significant in both the highFIS and expert cohorts, indicating that there was higher consensus in the attributes used to describe stimuli in the lowFIS cohort than in the highFIS or expert cohorts. Prior reports in the literature conflict on this point: some show that trained panelists use more terms to describe the product space (Chollet et al., 2011) while others show that experts tend to use

fewer words (Nestrud & Lawless, 2008). Our study suggests that both findings may be true. Here, we show experts do not use significantly more terms to describe the product space than a cohort of highFIS individuals who presumably lack formal culinary education, and that both the expert and highFIS cohorts use significantly more terms to describe the product space than the lowFIS cohort, who lack formal culinary education and have low food involvement. Again, it appears that type of expertise and education may play a significant role.

Another area of debate is the level to which trained panelists, experts, and untrained assessors are consensual in their descriptors. Some work suggests that trained panelists and experts tend to use more precise descriptions of samples where untrained panelists tend to use less specific terms (Clapperton & Piggott, 1979; Faye et al., 2004; Gains & Thomson, 1990; Gawel, 1997; Guerrero et al., 1997; Lawless, 1984), while other work suggests that chefs use fewer, more idiosyncratic words (Nestrud & Lawless, 2008).

In this study, in addition to the number of terms generated, the distribution of the generated attributes differed between the cohorts. The expert and highFIS cohorts were very similar in the distribution of the attributes that they generated, using roughly equal number of attributes to describe chemesthetic sensations, tastes/foods, and aromas/other sensations. On the other hand, in the lowFIS cohort, a majority of the terms generated related to the chemesthetic sensations, with slightly fewer relating to the taste/specific foods, and only about 15% of the terms relating to aromas/other sensations. Notably, the lowFIS cohort described stimuli in reference to one another (roughly 10% of the generated terms). This behavior was not seen in the highFIS or expert cohorts.

Overall, the highFIS and expert cohorts tended to behave similarly in the number of unique attributes that were generated per assessor, and the number of attributes that were significant in regression. While these two cohorts generated more than 1.5 times the number of unique attributes than the lowFIS group did, the low number of attributes recovered from regression suggests that there was more idiosyncratic behavior among these two cohorts, which both showed high scores on the FIS. Conflicting reports exist regarding whether experts and trained or untrained assessors use more or less descriptors and if the experts and trained assessors are more precise or more idiosyncratic than untrained assessors. A number of studies suggest that trained assessors and experts tend to be more precise in their descriptions, using words more efficiently, while untrained panelists tend to use more ambiguous terms (Chollet & Valentin, 2001; Clapperton & Piggott, 1979; Faye et al., 2004; Gains & Thomson, 1990; Gawel, 1997; Guerrero et al., 1997; Lawless, 1984). However, work by Nestrud and Lawless (Nestrud & Lawless, 2008) compared chefs to untrained assessors using a napping procedure to evaluate citrus juices and found that chefs tended to use fewer, unique terms and behaved more idiosyncratically when compared to consumers. The difference in these findings on word use may be due to differences in the type of training and expertise that the various experts and trained assessors in each of the studies or to the type of testing methodology employed in the studies. It is possible that the congruency of the results from this study with previously reported results may be influenced by the fact that Nestrud and Lawless (2008) collected their expert data at the same culinary school used here, using chefs with an average of 20 years of experience. Based on these results, it seems

possible that culinary education may not place the same emphasis on linguistic consensus as other types of formal training (e.g., wine expertise, descriptive panels). Accordingly, employing this methodology with individuals who have received formal culinary education from other well known institutions (e.g., Johnson & Wales University or Le Cordon Bleu) in the future may be informative. Additional data on the assessors' exposure and involvement with foods may provide a clearer understanding of why the lowFIS, highFIS, and expert cohorts differed in the ways observed here.

Previous work suggests the performance differences between experts and untrained assessors may be due to superior memory abilities of experts such that they are less impaired by delays between samples and may be able to better handle the increasing load on memory as sample number increases (Almeida et al., 1999; Chollet et al., 2005). The current study was not designed to explore these issues as they relate to differences between experts, trained panelists, and untrained assessors; rather we explored how differences in the level of experience may influence the way that assessors complete a free sorting task. Existing literature suggests that experts perform better on tasks such as perceptual mapping because they are able to describe their perception with more precise formal language, can discriminate better, and perceive more dimensions (Chollet et al., 2011; Roberts & Vickers, 1994; Solomon, 1990; Tang & Heymann, 2002; Torri et al., 2013). It has also been proposed that not just the kind of expertise but also the level of expertise with a sample type significantly influences product differentiation ability (Barcenas et al., 2004; Maitre, Symoneaux, Jourjon, & Mehinagic, 2010). Moreover, when compared to untrained assessors, experts and trained subjects tend to use non-hedonic criteria for sample differentiation (Delarue & Sieffermann, 2004). Roberts and Vickers (1994) found significant differences in the way that judges trained in the ADSA methods, trained panelists (i.e. descriptive analysis), and untrained assessors perceived cheeses: the training process for dairy judges resulted in a shift of these judges' focus as they were trained to focus on defects. Indeed, present results agree with prior findings, as the non-formally trained assessors with food involvement scores similar to those with formal training generated similar numbers of unique attributes, used similar numbers of groups during the sorting task, and generated the same number of clusters on the MDS map. The expert cohort did however appear to have less difficulty than the highFIS cohort, consistent with expectations as this cohort has the most focused education on food, and presumably expertise, of all cohorts in the present study.

5. CONCLUSION

Our results suggest type of expertise as well as type of experience (i.e. formal culinary education versus experiential learning) are important factors to consider when interpreting perceptual maps. While the highFIS group generally lacked formal culinary education, they tended to behave similarly to the formally educated chefs and culinary students in their use of descriptors. This suggests that greater Food Involvement may lead to experiential learning, which results in similarities between formally educated, or trained assessors and untrained assessors. The two groups with highFIS scores tended to be more descriptive in the number of terms that they used to describe samples, but there was also less consensus regarding these descriptors. A priori, we expected the culinary experts to have better

consensus in their descriptors (similar to wine experts) but during data collection it became clear that a number of the experts approached the task as an identification task, trying to identify the food or spice that the stimulus was derived from. The fact that many assessors in the expert cohort attended to the task differently from other cohorts could be the source for some of the variation observed in attribute consensus. Interestingly, while the highFIS cohort performed similarly to the expert cohort regarding attribute generation, this cohort performed similarly to the lowFIS cohort with regard to the configurations of the perceptual maps.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This work was supported by a National Institutes of Health grant from the National Institute of Deafness and Communication Disorders [DC010904] to J.E.H., United States Department of Agriculture Hatch Project PEN04332 funds, and funds from the Pennsylvania State University.

This manuscript was prepared in partial fulfillment of a Doctor of Philosophy degree at the Pennsylvania State University by N.K.B. The authors would like to thank Meghan Kane, Laura Boone, and Geneva Bonny for their help with data collection, and all of the participants at Penn State and the Culinary Institute of America for their participation in this study.

References

- Albin KC, Simons CT. Psychophysical evaluation of a sanshool derivative (alkylamide) and the elucidation of mechanisms subserving tingle. *PLoS One*. 2010; 5(3):e9520. [PubMed: 20209090]
- Almeida TC, Cubero E, O'mahony M. Same-Different Discrimination Tests With Interstimulus Delays Up To One Day. *Journal of Sensory Studies*. 1999; 14(1):1–18.
- Barcenas P, Elortondo FP, Albisu M. Projective mapping in sensory analysis of ewes milk cheeses: A study on consumers and trained panel performance. *Food Research International*. 2004; 37(7):723–729.
- Bell R, Marshall DW. The construct of food involvement in behavioral research: scale development and validation☆. *Appetite*. 2003; 40(3):235–244. [PubMed: 12798781]
- Bennett SM, Hayes JE. Differences in the chemesthetic subqualities of capsaicin, ibuprofen, and olive oil. *Chemical Senses*. 2012; 37(5):471–478. [PubMed: 22281531]
- Byrnes NK, Hayes JE. Personality factors predict spicy food liking and intake. *Food Quality and Preference*. 2013; 28(1):213–221. [PubMed: 23538555]
- Byrnes NK, Hayes JE. Gender differences in the influence of personality traits on spicy food liking and intake. *Food Quality and Preference*. 2015
- Byrnes NK, Nestrud MA, Hayes JE. Perceptual Mapping of Chemesthetic Stimuli in Naive Assessors. *Chemiosensory Perception*. 2015; 8(1):19–32. [PubMed: 26236421]
- Cartier R, Rytz A, Lecomte A, Poblete F, Krystlik J, Belin E, et al. Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*. 2006; 17(7):562–571.
- Chollet S, Lelièvre M, Abdi H, Valentin D. Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference*. 2011; 22(6):507–520.
- Chollet S, Valentin D. Impact of training on beer flavor perception and description: are trained and untrained subjects really different? *Journal of Sensory Studies*. 2001; 16(6):601–618.

- Chollet S, Valentin D, Abdi H. Do trained assessors generalize their knowledge to new stimuli? *Food Quality and Preference*. 2005; 16(1):13–23.
- Clapperton J, Piggott J. Flavour characterization by trained and untrained assessors. *Journal of the Institute of Brewing*. 1979; 85(5):275–277.
- Cliff M, Heymann H. Descriptive analysis of oral pungency. *Journal of Sensory Studies*. 1992; 7:12.
- Delarue J, Sieffermann JM. Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*. 2004; 15(4):383–392.
- DeNeve KM, Cooper H. The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychol Bull*. 1998; 124(2):197–229. [PubMed: 9747186]
- Faye P, Brémaud D, Daubin MD, Courcoux P, Giboreau A, Nicod H. Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality and Preference*. 2004; 15(7):781–791.
- Faye P, Brémaud D, Teillet E, Courcoux P, Giboreau A, Nicod H. An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*. 2006; 17(7):604–614.
- Gains N, Thomson DM. Sensory profiling of canned lager beers using consumers in their own homes. *Food Quality and Preference*. 1990; 2(1):39–47.
- Gawel R. The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies*. 1997; 12(4):267–284.
- Giacalone D, Ribeiro LM, Frøst MB. Consumer-based product profiling: application of partial napping® for sensory characterization of specialty beers by novices and experts. *Journal of Food Products Marketing*. 2013; 19(3):201–218.
- Gilmore MM, Green BG. Sensory irritation and taste produced by NaCl and citric acid: effects of capsaicin desensitization. *Chemical Senses*. 1993; 18(3):257–272.
- Green BG. Temperature perception and nociception. *Journal of neurobiology*. 2004; 61(1):13–29. [PubMed: 15362150]
- Green BG, Hayes JE. Capsaicin as a probe of the relationship between bitter taste and chemesthesis. *Physiology & Behavior*. 2003; 79(4):811–821. [PubMed: 12954427]
- Green BG, Hayes JE. Individual differences in perception of bitterness from capsaicin, piperine and zingerone. *Chemical Senses*. 2004; 29(1):53–60. [PubMed: 14752040]
- Guerrero L, Gou P, Arnau J. Descriptive Analysis Of Toasted Almonds: A Comparison Between Expert And Semi-Trained Assessors. *Journal of Sensory Studies*. 1997; 12(1):39–54.
- Hayes JE, Pickering GJ. Wine expertise predicts taste phenotype. *American journal of enology and viticulture*. 2011 ajev. 2011.11050.
- Husson F, Lê S, Cadoret M. *SensMineR: Sensory data analysis with R*. R package version 1.20. 2014
- Kennedy J, Heymann H. Projective mapping and descriptive analysis of milk and dark chocolates. *Journal of Sensory Studies*. 2009; 24(2):220–233.
- Krzanowski, WJ.; Marriott, FHC. *Multivariate analysis*. Edward Arnold; London: 1994.
- Lawless HT. Flavor description of white wine by “expert” and nonexpert wine consumers. *Journal of Food Science*. 1984; 49(1):120–123.
- Lawless HT, Glatter S. Consistency of multidimensional scaling models derived from odor sorting. *Journal of Sensory Studies*. 1990; 5(4):217–230.
- Lawless, HT.; Horne, J. *Category Reviews and Multidimensional Scaling*. C. University; Ithaca, NY: 2000.
- Maitre I, Symoneaux R, Jourjon F, Mehinagic E. Sensory typicality of wines: How scientists have recently dealt with this subject. *Food Quality and Preference*. 2010; 21(7):726–731.
- Marshall D, Bell R. Relating the food involvement scale to demographic variables, food choice and other constructs. *Food Quality and Preference*. 2004; 15(7):871–879.
- Nestrud, MA.; Lawless, HT. The distribution of the RV coefficient for comparing multivariate configurations. In: Guelph, editor. *Poster. Sensometrics*. Canada: Elsevier; 2008.
- Nestrud MA, Lawless HT. Perceptual mapping of apples and cheeses using projective mapping and sorting. *Journal of Sensory Studies*. 2010; 25(3):390–405.

- Nestrud MA, Lawless HT. Recovery of subsampled dimensions and configurations derived from napping data by MFA and MDS. *Attention, Perception, & Psychophysics*. 2011; 73(4):1266–1278.
- Pagès J. Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference*. 2005; 16(7):642–649.
- Parr WV, White KG, Heatherbell DA. Exploring the nature of wine expertise: what underlies wine experts' olfactory recognition memory advantage? *Food Quality and Preference*. 2004; 15(5):411–420.
- Perrin L, Symoneaux R, Maître I, Asselin C, Jourjon F, Pagès J. Comparison of three sensory methods for use with the Napping® procedure: Case of ten wines from Loire valley. *Food Quality and Preference*. 2008; 19(1):1–11.
- Risvik E, McEwan JA, Rødbotten M. Evaluation of sensory profiling and projective mapping data. *Food Quality and Preference*. 1997; 8(1):63–71.
- Robert P, Escoufier Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied statistics*. 1976:257–265.
- Roberts AK, Vickers ZM. A Comparison Of Trained And Untrained Judges' evaluation Of Sensory Attribute Intensities And Liking Of Cheddar Cheeses. *Journal of Sensory Studies*. 1994; 9(1):1–20.
- Schiffman, SS.; Reynolds, ML.; Young, FW.; Carroll, JD. Introduction to multidimensional scaling: Theory, methods, and applications. Academic press; New York: 1981.
- Solomon GEA. Psychology of novice and expert wine talk. *The American Journal of Psychology*. 1990:495–517.
- Tang C, Heymann H. Multidimensional Sorting, Similarity Scaling And Free-Choice Profiling Of Grape Jellies. *Journal of Sensory Studies*. 2002; 17(6):493–509.
- Torri L, Dinnella C, Recchia A, Naes T, Tuorila H, Monteleone E. Projective mapping for interpreting wine aroma differences as perceived by naïve and experienced assessors. *Food Quality and Preference*. 2013; 29(1):6–15.

- Experience, either formal culinary training or experiential learning, significantly influences lexical development.
- Lexical development influences the outcome of a sorting task with a descriptive portion.
- Individuals without formal culinary training, both high and low Food Involvement cohorts, generated significantly similar perceptual maps.
- Naïve individuals with high Food Involvement Scale scores acted as an intermediate between naïve low-FIS individuals and experts.

Figure 1.
Distribution of FIS scores between the three cohorts in this study.

Figure 2.

Perceptual map of 11 compounds sorted in a free sorting task by 26 assessors with low Food Involvement Scale scores. Regression was performed to regress descriptors generated by participants onto the perceptual map. Stimuli include allyl isothiocyanate (AITC), capsaicin (CAP), carvacrol (CARV), cinnamaldehyde (CINN), citric acid (CA), eucalyptol (EUCA), eugenol (EUG), huajiao (HJ), menthol (MEN), quinine (Q), and zingerone (ZING).

Figure 3.
Two-dimensional perceptual map similar to Figure 2, except participants were from the highFIS score group ($n = 25$).

Figure 4. Two-dimensional perceptual map similar to Figures 2 and 3, but for the expert cohort (i.e. those with formal culinary education; n = 32).

Table 1

Mean number of attributes generated and mean number of groups formed by each cohort.

Cohort	Mean number of attributes (+/- SE)	Mean number of groups formed (+/- SE)
LowFIS	7.69 ± 0.58 ^a	5.5 ± 0.4 ^a
HighFIS	9.04 ± 0.66 ^a	6.3 ± 0.2 ^a
Experts	11.80 ± 0.71 ^b	6.2 ± 0.3 ^a

Superscript letters indicate statistically significantly different values ($p < 0.05$).

Table 2

Description of how three cohorts used descriptors differently.

Cohort	Touch sensations	Tastes/foods	Aromas/non-food sensations
LowFIS	15/35 (43%)	11/35 (31%)	5/35 (14%)
HighFIS	18/56 (32%)	26/56 (46%)	12/56 (22%)
Experts	19/54 (35%)	24/54 (44%)	11/54 (21%)