

Invited Review: Sensory Analysis of Dairy Foods

M. A. Drake

Department of Food Science, Southeast Dairy Foods Research Center, North Carolina State University, Raleigh 27695-7624

ABSTRACT

Sensory quality is the ultimate measure of product quality and success. Sensory analysis comprises a variety of powerful and sensitive tools to measure human responses to foods and other products. Selection of the appropriate test, test conditions, and data analysis result in reproducible, powerful, and relevant results. Appropriate application of these tests enables specific product and consumer insights and interpretation of volatile compound analyses to flavor perception. Trained-panel results differ from dairy judging and grading and one objective of this review is to clearly address and demonstrate the differences. Information on available sensory tests, when and how to use them, and the powerful results that can be obtained is presented.

Key words: sensory analysis, sensory methods, flavor, texture

INTRODUCTION

When dealing with dairy foods, sensory quality is always involved on some level. The best raw ingredients make the best finished products so sensory quality is a critical aspect of dried dairy ingredients and fluid milk. Sensory perception of finished products such as ice cream and cheese is also crucial. Sensory perception is one of the keys to the widespread flavorful and wholesome image that dairy foods continue to enjoy with the consumer. Due to the integral nature of sensory perception with dairy foods, a sensory measurement is often the final step in many experiments or applications. In some cases, a gross measurement of product quality or consistency may be all that is required; however, for most product and market research, more detailed and complex information on sensory properties is needed.

Sensory science is a relatively young discipline and has been in existence for roughly 60 yr. Many attribute

the conception of sensory science to the 1940s with the development of consumer or hedonic food acceptance methodologies by the US Army Corps of Engineers (Peryam and Pilgrim, 1957). However, in reality, sensory science can be traced back to the 1800s, with the development of psychological theories to measure and predict human responses to external stimuli (Lawless and Heymann, 1999a). Certainly, the importance of sensory quality is ageless, with basic capitalism driving individuals to market and sell the best and freshest products because these products demand the top dollar. As with other fields of science, sensory science has progressed with time and continues to evolve; thus, current research should incorporate the most progressive tools available that are suited to the purpose and goals of the research.

Specific scientific methods have been developed to accurately, reproducibly, and objectively measure or estimate human responses to stimuli. Like and dislike are not the only questions that are answered by sensory analysis. Consumer perception and emotional responses can also be addressed (Young et al., 2004); the impact of storage (Park et al., 2006), ingredient substitution (Childs et al., 2007), and package and process variability can be determined (Carunchia Whetstine et al., 2006b, 2007); and relationships can be established between instrumental tests and sensory perception (Avsar et al., 2004; Wright et al., 2006). In fact, dozens of types of sensory tests exist and can be fine-tuned to meet a specific objective. Too often, sensory testing is an afterthought to an experiment and an inappropriate test is used or the appropriate test is selected but misused. When these situations occur, sound results cannot be collected and correct conclusions cannot be made, as with any other scientific test inappropriately selected or conducted. If questions about the sensory properties of a food need to be answered, then the sensory method(s) and experimental design should be selected before the data are collected. Knowledge of the tests available, when and how to use them, and the powerful results that can be obtained are presented in this review.

TYPES OF SENSORY TESTS

Traditional Sensory Tests

Sensory analysis can be categorized into 3 basic categories or groups of tests. The first group of tests includes

Received May 2, 2007.

Accepted June 25, 2007.

E-mail: mdrake@unity.ncsu.edu

traditional tools such as USDA grading and ADSA scorecard judging. These sensory tests were developed in the early 1900s by the dairy industry to ensure product quality and consistency and to encourage and train students (Bodyfelt et al., 1988). By these techniques, a product is assigned an overall quality score or grade based on a designated list of defects. When initially developed, these tools played a role in product research but with the advent of modern sensory tests, they are only used today to troubleshoot product quality problems, to train students, and to ensure baseline quality of government commodities. These methods are fast and practical in a large manufacturing facility or quality control environment when decisions on product quality must be made quickly. An experienced individual can rapidly identify gross product quality defects, their potential sources, and take corrective or preventative action(s). However, scorecard judging and grading methods suffer from several scientific shortcomings that make them entirely unsuitable for use in product or market research. Many of the terms for quality and grade defects are outdated and not well defined, making identification and scoring subjective at best. The quality/grade defects are not entirely descriptive of the products. This means that 2 products can receive the same quality scores—and even the same quality defects and specific point deductions—and still be quite different in flavor or texture (Drake, 2004).

Quality downgrades for defects are not consistent (different defects result in different point deductions) and within a specific defect, the point deductions are not consistent. These point deductions ultimately mean that score assignment is not linear, which precludes the use of powerful parametric statistics, the standard statistical analysis with other sensory and instrumental tests. Finally, quality perception by judges or graders is not associated with consumer acceptance or preference, and quality results do not describe the total sensory profile of the product. These issues have been reviewed in detail (Singh et al., 2003; Delahunty and Drake, 2004; Bodyfelt et al., accepted). These tools should not be used in research for any reason. There are numerous mainstream sensory tools (from very simple to complex) based on the psychological, physical, and physiological science of human responses to external stimuli; that is, sensory science, that can readily be applied to meet any specific sensory research objective in dairy foods, and only these mainstream tools should be used in current research.

Mainstream Sensory Tests

Analytical Sensory Tests: Discrimination Tests. Mainstream sensory tests (those tests that are estab-

lished and scientifically recognized sensory tests rather than commodity-based traditional sensory tests) include 2 basic groups of tests: analytical tests and affective or consumer tests. Within each of these categories are groups of sensory tools for specific objectives and within each of these groups are several different specific tests or approaches. The best-known analytical sensory test is the difference or discrimination test. The sole objective of a difference test in its simplest and most-used form is to determine if a difference exists between 2 or more products (Meilgaard et al., 2007a). The most widely used types of difference tests are the triangle and duo-trio tests, although there are several others. Selection of the appropriate difference test to use is often determined by amount of sample, number of samples, testing conditions, nature of the potential difference (e.g., known or unknown) and specific test objectives. These tests are easy to set up and administer and the results are easily determined using a simple binomial calculation or published tables (Lawless and Heymann, 1999b). The number of panelists required varies depending on the specific goal; generally, 25 to 50 panelists are recommended (Lawless and Heymann, 1999b). It is important to keep in mind that the sole purpose of these tests is to determine if a difference exists. Difference tests are one of the most commonly misused sensory tools because the nature of the difference, the degree of difference, or consumer preference cannot be determined using this test nor can these questions be asked of panelists when taking a difference test. If these questions need to be answered, a different or additional sensory test is required.

Directional difference tests are situations where the nature of the difference tested is known (e.g., chocolate milks in which sweetener concentration is the only difference). In this case, a more sensitive difference test, the paired comparison, 2-alternative forced choice (AFC) or 3-AFC test can be used (Meilgaard et al., 2007b). These tests are similar in set-up to duo-trio and triangle tests but the nature of the difference is specified to the subjects. These tests can also be applied to traditional difference testing in which the nature of the difference is not known, but each panelist determines their own perception of the difference in a series of warm-up samples provided. Studies have indicated that, when applied in this method, these tests are more sensitive to difference detection than are triangle and duo-trio tests (Jiamyangyuen et al., 2002). Other subcategories of difference tests such as a degree of difference (DOD) test can be used to quantify the degree of difference among samples or specific attribute differences among samples (Meilgaard et al., 2007b), but this more advanced test generally requires fewer but more experienced or trained panelists. In contrast, another

type of test similar to, but distinct from, difference testing is similarity testing; this test is set up similar to a difference test, but generally requires larger numbers of panelists (>75; Meilgaard et al., 2007a). There are numerous types of difference tests and numerous powerful applications of these tests beyond “simple” difference detection.

Threshold Tests. Threshold tests are another category of analytical sensory tests with a specific function: to determine thresholds. A threshold is defined as the lowest concentration at which a sensory response is detectable (Lawless and Heymann, 1999c; Meilgaard et al., 2007c). There are other types of thresholds such as absolute threshold (previously defined), recognition threshold (lowest concentration at which a compound can be recognized), difference threshold (concentration at which differences in stimuli can be detected), terminal threshold (concentration above which there is no perceived increase in sensory stimulus), orthonasal threshold (threshold of volatile compound orthonasally), and retronasal threshold (threshold of volatile compound retronasally). The latter are determined by having subjects wear noseclips when taking a mouthful of the sample followed by removal of the noseclip once the compound is in the mouth. Thresholds are often applied to undesirable (off-flavor compounds) as well as desirable components in foods. For example, at what concentration is dimethyl trisulfide an off-flavor in whey protein isolate (WPI)? Such a question can only be answered by threshold testing of dimethyl trisulfide in water and WPI to determine what concentrations are detected by humans (Wright et al., 2006). A desirable flavor example might be what concentration of diacetyl is detected by consumers in cottage cheese? Again, a threshold test will be needed to determine the threshold concentration of diacetyl in cottage cheese. Of course, the previous example assumes that previous consumer sensory tests (discussed later) have established that diacetyl is a desirable attribute in cottage cheese. Thresholds can thus provide a powerful tool in relating sensory perception to instrumental analysis of volatile and nonvolatile compounds (Robinson et al., 2004, 2005; Carunchia Whetstone et al., 2005).

Searching the scientific literature for the threshold of a compound can be challenging. Indeed, threshold ranges of more than 1,000-fold for many compounds are reported throughout the literature (Rychlik et al., 1998; van Gemart, 2003). Several issues must be addressed to determine an accurate and reproducible threshold value. Thresholds are affected by several challenges and perhaps the largest impact is proper and consistent testing procedure. This includes an appropriate threshold test method, an appropriate number of panelists, and consistent methodology. The 2

most common threshold procedures are the ascending forced choice or 3-AFC method (method of limits) and the R-index method (signal detection method; ASTM, 1992; Lawless and Heymann, 1999c). An appropriate number of panelists is the next key to a sound threshold. A large number of individuals need to be tested to obtain a reliable threshold. Seventy-five to 100 is often considered a sound number of individuals although testing 30 to 40 individuals on multiple days can also approach a sound best estimate threshold. Thresholds obtained from fewer individuals should be suspect.

Finally, testing procedure can also be a source of threshold variability. Temperature and headspace will both influence threshold, and the nature of the influence can be variable (increase or decrease; Meilgaard et al., 2007c). Generally, lower temperatures and smaller headspace volumes will increase thresholds. The chemical composition of the food matrix also plays a role. For simplicity, water or oil are generally used as matrices for threshold testing. Food is much more complex with macrocomponents (fat, protein, carbohydrates) that can bind volatile compounds as well as other inherent volatile components of the food itself that can influence threshold of the compound being tested. Karagül-Yüceer et al. (2004) conducted threshold testing on selected volatile compounds in water and in a relatively simple food matrix: skim milk. They reported that orthonasal thresholds for some compounds were unchanged between water and skim milk, but for others, threshold in skim milk was increased more than 100-fold.

Descriptive Sensory Analysis. The final general group of analytical sensory tests is descriptive analysis. Descriptive analysis consists of training a group of individuals (generally 6 to 12) to identify and quantify specific sensory attributes or all of the sensory attributes of a food. This sensory tool, unlike the previous analytical tests, which use untrained or instructed/screened individuals, requires training of the panelists. The extent of the training is dependent on the complexity of the sensory attributes that will be profiled. Training may be as brief as a few hours if there are only a few attributes and they are distinct in the samples. On the other hand, a significant time or financial commitment is required if flavor profiling of 16 attributes (or more) of Cheddar cheese is desired. The power of descriptive analysis is that the panel and its training can be adjusted to meet the specific project goals. The panel can be trained on a few attributes or a large number of attributes.

The panelists are trained (sometimes for several hundred hours) to operate in unison as an instrument, and each individual panelist serves a function analogous to an individual sensor on an instrument. Panel and

panelist performance should be monitored throughout training to identify problem areas and to track discriminatory ability of the panel. There is no set rule for what constitutes optimal panel or panelist performance. Instead, this is up to the judgment of the panel leader and their knowledge of the panel and panelists as well as the products and attributes. If time for training is short or the product is fatiguing or heterogeneous, increasing panel replications can further enhance panel discriminating power. Chambers et al. (2004) trained panelists for 4, 60, or 120 h on tomato sauce attributes. Their results confirmed that minimal training was required for some texture and flavor attributes, whereas extensive training was required to discriminate other attributes. If the specific nature of the product differences and number of panelists are known, specific panel target performance (standard deviation) can be computed and modeled (Gacula and Rutenbeck, 2006). Generally, this information is not known and the panel leader must rely on data analysis collected from preliminary training sessions to confirm panel performance. The panel replicates measurements analogous to replication of instrumental measurements and the data collected are analogous to instrumental data. There are different approaches and training techniques for descriptive sensory analysis but the primary goal of these different approaches is the same: a powerful instrument to document sensory properties. Techniques and approaches to descriptive analysis are reviewed elsewhere (Lawless and Heymann, 1999d; Murray et al., 2001; Delahunty and Drake, 2004).

A trained sensory panel should produce results analogous to instrumental data. As such, the sensory instrument (panel) should be as precise and reproducible as possible. Training with defined sensory languages and replication of panel measurements are used to achieve this goal. One way of minimizing variability is panel training, in which panelists are presented with the sensory language (or lexicon) and then discuss these attributes as they relate to the products that will be evaluated. However, a crucial step to facilitate panel training and panel performance and to establish any relationship to physical or instrumental measurements is to have clearly defined terms for sensory attributes (Drake and Civille, 2003). Defined terms facilitate panel training and minimize variability but they also set the parameters for understanding instrumental measurement of the sensory attribute. For example, is cheese firmness measured by compression with fingers, bite force with incisors, bite force with molars, or compression between the tongue and the hard palate? Is free fatty acid flavor defined as the aroma or flavor reminiscent of hexanoic acid, butyric acid, methyl octanoic acid, or any free fatty acid? In the case of texture, what

is the defined size and shape of the sample? Cheese firmness might be measured by the fingers, tongue, incisors, or molars depending on the type of cheese. Product temperature may influence attribute (flavor and texture) perception and should be reported.

Ideally, references (food or chemical examples) are also provided in addition to attribute definitions to aid panelists in training and attribute identification and scale usage. These are not necessary but certainly will assist in replication or comparison of results across multiple laboratories. Delahunty and Drake (2004) reviewed sensory lexicons for cheese flavor and texture. Foegeding and Drake (2007) later reviewed sensory and instrumental perception of cheese texture. It is important to keep in mind that references serve to focus the panel on the general concept of the specified attribute. They are not necessarily identical to the attribute. As such, multiple attributes often work best because different panelists associate better with different references. Ideally, food references for attributes should be compositionally defined (moisture, fat, pH); however, this is not always possible with unprocessed food references nor absolutely necessary (Drake and Civille, 2003).

Table 1 demonstrates a published lexicon for Cheddar cheese flavor with definitions and references and adaptation of this language to a different type of cheese (Swiss; Table 2). Similar languages have been identified for dried dairy ingredients, chocolate milk, and butter (Brown et al., 2003; Drake et al., 2003; Thompson et al., 2004; Krause et al., 2007). As previously addressed, the sensory language can be simple or complex depending on the specific test objective(s). Further, sensory languages can expand and are modified and clarified with time and usage.

Affective (Consumer) Tests. The third group of sensory tests is affective or consumer tests. Like analytical sensory tests, a large array of specific and sensitive tools fit in this category. To the beginner or the uninformed, this group of tests is simply measuring preference and liking. However, this group of tests is actually expansive, diverse, and very complex. Qualitative and quantitative tests are available. Consumer tests involve testing with consumers. This issue may seem obvious, but by their very objective, these tests mean that trained panelists *should not* be used. Once individuals are trained to identify and quantify attributes of a product(s), they cease to be typical consumers, because the typical consumer is not trained and does not sequentially identify and quantify specific attributes in their foods. As such, a trained panelist is now more informed, aware, and potentially more critical and sensitive than the average or typical consumer. Furthermore, when quantitative consumer tests are conducted, their objec-

Table 1. Basic Cheddar cheese flavor language

Descriptor	Definition	Reference
Cooked/milky	Aromatics associated with cooked milk	Pasteurized skim milk heated to 85°C for 30 min
Whey	Aromatics associated with Cheddar cheese whey	Fresh Cheddar whey
Diacetyl	Aromatic associated with diacetyl	Diacetyl, 20 ppm
Milkfat/lactone	Aromatics associated with milkfat	Fresh coconut meat, heavy cream, δ -dodecalactone, 40 ppm
Fruity	Aromatics associated with different fruits	Fresh pineapple, ethyl hexanoate, 40 ppm
Sulfur	Aromatics associated with sulfurous compounds	Boiled egg, H ₂ S bubbled through water, freshly struck match
Free fatty acid	Aromatics associated with short-chain fatty acids	Butyric acid, 20 parts per thousand
Brothy	Aromatics associated with boiled meat or vegetable soup stock	Canned potatoes, Wyler's low sodium beef broth cubes, methional, 20 ppm, 2-methyl-3-furanthiol
Nutty	The sweet roasted aromatic associated with various nuts, wheat germ, unsalted wheat thins	Lightly toasted unsalted nuts, roasted peanut oil, 2/3 methyl butanal, 500 ppm
Catty	Aroma associated with tom-cat urine	2 mercapto-2 methyl-pentan-4 one, 20 ppm
Cow/barny	Aroma associated with barns and animal sheds, reminiscent of ruminant sweat and urine	Mixture of isovaleric acid and p-cresol, 100 ppm
Mothball/feed	Aroma associated with mothballs or protein catabolism, sometimes reminiscent of silage or grass compost	Mothballs, indole or skatole, 50 ppm
Sour	Fundamental taste sensation elicited by acids	Citric acid (0.08 g/100 mL in water)
Salty	Fundamental taste sensation elicited by salts	Sodium chloride (0.5 g/100 mL in water)
Sweet	Fundamental taste sensation elicited by sugars	Sucrose (5 g/100 mL in water)
Bitter	Fundamental taste sensation elicited by caffeine or quinine	Caffeine (0.08 g/100 mL in water)
Umami	Chemical feeling factor elicited by certain peptides and nucleotides	Monosodium glutamate (1 g/100 mL in water)

¹Expanded from Drake et al. (2001) with permission from Wiley Blackwell.

tive is to determine or infer consumer likes and dislikes. Consumers are highly variable and constantly changing due to age, advertising, new experiences, new products, and so on. For this reason, large and successful companies have large sensory or market research departments that conduct these tests regularly and with large numbers of representative consumers. Demographic information (age, gender, product usage rate) is generally collected from consumers to determine if these variables influence product liking. Additional information (income, ethnicity, product perceptions/attitudes) can also be probed in the screener if desired. For this reason, these screeners are sometimes called usage and attitude screeners (or U&A information). Even for small research projects or objectives, a minimum of 50 consumers is recommended to make any conclusion about product liking or preference, and these should be product consumers, not trained panelists (IFT/SED, 1981; Resurreccion, 1998; Hough et al., 2006; Meilgaard et al., 2007d). The reader is referred to several textbooks that address these issues in detail (Resurreccion, 1998; Lawless and Heymann, 1999e; Meilgaard et al., 2007d; Moskowitz et al., 2007).

Quantitative Consumer Tests. Preference and acceptance tests are the most widely known group of quantitative consumer tests (Lawless and Heymann, 1999e; Meilgaard et al., 2007d). Preference and acceptance testing are often used interchangeably but they are distinct tests. In preference testing, consumers are presented with 2 or more samples and asked to indicate

which sample they prefer. If more than 2 samples are presented, consumers can also rank their preference (preference ranking). The test is generally forced choice; that is, a preference must be indicated. A preference test is easy to conduct and the question is readily understood by consumers of all ages. Nonparametric statistical analysis can be applied to determine differences. However, a primary drawback is that degree of liking is not determined. Consumers can dislike products and still prefer one when forced to choose. Furthermore, other consumer questions besides overall liking can be asked with acceptance testing and preference can be inferred from acceptance testing. In short, with acceptance testing, more information along with preference can be obtained.

Acceptance testing is also called degree of liking. Consumers are presented with products and asked to indicate degree of liking on a scale. The most commonly used scale is the 9-point hedonic scale. This scale is bipolar (the anchors are dislike and like) and has been widely used since its invention in the 1940s (Peryam and Pilgrim, 1957; Schutz and Cardello, 2001). In this sense, it has certainly stood the test of time. The scale can be presented numerically or verbally, horizontally or vertically (Schutz and Cardello, 2001) and is used to indicate differences in consumer liking of products. Other adaptations of this scale include a 7-point scale and a smiley face scale that can be used with children or those that do not speak or read English. Research has suggested that issues of central tendency and unequal

Table 2. Swiss cheese descriptive analysis lexicon adapted and modified from the Cheddar cheese lexicon (Drake et al., 2001; Liggett et al., 2007)

Descriptor	Definition	Reference
Cooked/milky	Aromatics associated with cooked milk	Skim milk heated to 85°C for 30 min
Whey	Aromatics associated with Cheddar cheese whey	Fresh Cheddar whey
Diacetyl	Aromatic associated with diacetyl	Diacetyl
Milkfat	Aromatics associated with milkfat	Fresh coconut meat, heavy cream, δ -dodecalactone
Vinegar	Aromatics associated with vinegar	Distilled white vinegar, acetic acid
Dried fruit	Aromatics associated with dried fruits, specifically peaches and apricots	Dried apricot half
Fruity	Aromatics associated with different fruits	Fresh pineapple, ethyl hexanoate
Sulfur/eggy	Aromatics associated with cooked eggs	Hardboiled egg, mashed
Sulfur/cabbage	Aromatics associated with cooked cabbage	Boiled cabbage, dimethyl trisulfide
Cheesy/butyric acid	Aromatics associated with butyric acid	Butyric acid
Brothy	Aromatics associated with boiled meat or vegetable stock	Canned potatoes, Wyler's low sodium beef broth cubes, methional
Nutty	The nut-like aromatic associated with different nuts	Lightly toasted unsalted nuts, unsalted cashew nuts, unsalted wheat thins
Sweaty	Aromatic associated with human sweat	Hexanoic acid
Cowy/phenolic	Aromas associated with barns and stock trailers, indicative of animal sweat and waste	Band-aids, p-cresol, phenol
Sour	Fundamental taste sensation elicited by acids	Citric acid (0.08% in water)
Bitter	Fundamental taste sensation elicited by various compounds	Caffeine (0.08% in water)
Salty	Fundamental taste sensation elicited by salts	Sodium chloride (0.5% in water)
Sweet	Fundamental taste sensation elicited by sugars	Sucrose (5% in water)
Umami	Chemical feeling factor elicited by certain peptides and nucleotides	Monosodium glutamate (1% in water)
Prickle	Chemical feeling factor of which the sensation of carbonation on the tongue is typical	Soda water
Metallic	Chemical feeling factor elicited by metallic objects in the mouth	Aluminum foil

scale intervals are shortcomings of this scale, and other scales such as labeled affective magnitude (**LAM**) scales have been proposed as more sensitive alternatives (Schutz and Cardello, 2001; Greene et al., 2006). More recent research has suggested that liking and disliking are actually completely different thought processes and should not be scaled on the same continuum (Herr and Page, 2004).

The just-about-right (JAR) scale is another often-used scale that is a subcategory of acceptance testing (Lawless and Heymann, 1999e). This test is often used in product development or optimization studies because the experimenter can probe if a specific product attribute (such as sweetness or chocolate flavor) is "just about right." There are a limited number of categories and only nonparametric statistical analysis is appropriate. Attribute intensities can also be scaled by consumers using 9-point intensity scales although results should be interpreted with caution because consumers are best at addressing likes and dislikes, and studies have suggested that asking consumers other questions besides overall liking can influence their liking responses (Popper et al., 2004). The overall liking question should always be asked first. Finally, trained panelists are best for scaling attribute intensities.

The 9-point hedonic scale will certainly continue to be a mainstream quantitative consumer research tool.

Indeed, although studies have suggested that the LAM scale or nonpolar like and dislike scales may be more sensitive in certain situations, in these studies the 9-point hedonic scale was still a robust and perhaps more conservative estimate of consumer liking (Schutz and Cardello, 2001; Greene et al., 2006). As with any sensory test, it is important to remember that specific situations may call for a more specialized scale than the traditional 9-point hedonic scale. Such situations might include testing with children, non-English-speaking populations, or consumers in different countries (Kroll, 1990; Lawless and Heymann, 1999e). For most situations, or the standard research project where the goal is simply to determine if differences exist between products in consumer acceptance, the 9-point hedonic scale is the scale of choice.

Conjoint analysis is another group of quantitative consumer tests that can be used to probe consumer perceptions. Unlike preference and acceptance tests, which generally deal with actual products that are tasted/evaluated by the consumer, conjoint analysis does not require actual products. Conjoint or trade-off analysis is a technique that takes into account the fact that consumers make choices or trade-offs between independent (yet conjoined) attributes in a product when making a purchase decision (Orme, 2006). Consumers are presented with product attributes and are then

asked to go through a series of trade-offs. Quantitative data are generated that can be subjected to traditional statistical analyses. The goal is determination of which product attribute(s) are most important to the consumer, without having to manufacture prototypes. For example, Jones et al. (accepted) and Childs et al. (accepted) used conjoint analysis to determine which aspects of meal-replacement bars and beverages were most crucial to consumer selection and purchase and if specific protein type (whey or soy) had any effect on these issues. Preference mapping (discussed in more detail later) is another quantitative approach to understanding consumer likes and dislikes (Schlich, 1995).

Qualitative Consumer Tests. The final group of consumer research tools is qualitative tools. Using these tools, insights into consumer perceptions, needs, and desires can be probed for product development, advertising, and development of quantitative screeners and questionnaires. The primary tests in this group are the focus group and the interview. Focus groups are an example of a qualitative research tool in which an experienced moderator leads a group of 8 to 12 participants through a guided discussion. The conversation typically lasts for 1.5 to 2 h. The session is tape-recorded or videotaped, or external individuals may observe the session and record common themes. Subjective information about product attributes, preferences, and motivations can be gained in this manner (Lawless and Heymann, 1999f; Krueger and Casey, 2000; Meilgaard et al., 2007d). This test is widely used in market research. Focus groups have been used in various food studies examining a number of issues including food preference, safety, and usage (Sherlock and Labuzza, 1992; Cotunga and Vickery, 2004; Kosa et al., 2004). Optimally, a focus group is conducted in triplicate with the target sampling of consumers. Common themes and consensus opinions should be consistent among the 3 groups (similar to replications) for the results to be considered sound or valid (Krueger and Casey, 2000). The interview tool is conducted similarly except that it is generally a one-on-one exercise. Although more time-consuming, additional personal or detailed information may be obtained by using the interview tool. Because these tools are qualitative in nature and generally, low numbers of consumers are polled, results must be interpreted with caution. Ideally, a quantitative test would be conducted as a follow-up to confirm or expand findings.

SENSORY APPLICATIONS

Trained panel results differ from dairy judging and a relevant objective of this review is to clearly address and demonstrate the differences. A good example of

this issue was presented by Rehman et al. (accepted). Cheddar cheeses were manufactured using different starter cultures and evaluated by a trained sensory panel and industry graders. No consistent differences were identified by graders. In contrast, several specific flavor differences were characterized by the trained panel. The reasons for lack of difference from graders were 2-fold: 1) inconsistent use of the scorecard by the different graders and 2) lack of available relevant categories to score flavor differences on the grading scorecard. The differences between grading and descriptive sensory analysis were demonstrated with a very simple product such as skim milk powder in Figure 1. Both skim milk powders received the same grade (US Extra) and had the same "defect" of cooked. Thus, judging or grading would not differentiate these products and potentially valuable information would be overlooked. A similar example with Cheddar cheese was provided by Drake (2004).

The trained descriptive sensory panel is a qualitative and quantitative instrument to document sensory properties of foods. Principal component analysis (PCA) is a multivariate data compression technique that allows multiple treatments to be graphically displayed as they are differentiated by multiple variables. As such, this technique is often applied to assess how several products were differentiated by several sensory descriptors. The principal components are linear combinations of the variables that explain a certain amount of the variability within the data set. Figures 2 and 3 demonstrate use of a trained panel with a defined sensory language to document differences in whey protein flavor and cheese texture. These data demonstrate how descriptive analysis can be applied to understand flavor variability within a product category, whey protein (Figure 2), or to evaluate the effect of fat reduction on texture of cheese (Figure 3). From Figure 2, sensory variability is observed among whey protein concentrate 80% (WPC80) and WPI; in general, WPI are grouped as are WPC80. Outliers WPI5, WPI7, and WPC7 would warrant a closer look at the sensory means to precisely characterize differences, although the biplot suggests that they are differentiated by bitter taste (WPC7) and soapy flavors (WPI5, WPI7). Similarly, Figure 3 demonstrates that fat reduction in Gouda cheese resulted in increased springiness (hspring) and rate of recovery (hrate of recov), regardless of cheese age because cheeses 1, 16, 17, and 18 are fat-reduced cheeses. Cheeses 16 and 17 are more fat reduced than cheeses 1 and 18 and display decreased breakdown, smoothness, adhesiveness, and mouthcoating compared with cheeses 1 and 18.

Appropriate sensory techniques can be used to enhance product understanding, establish relationships

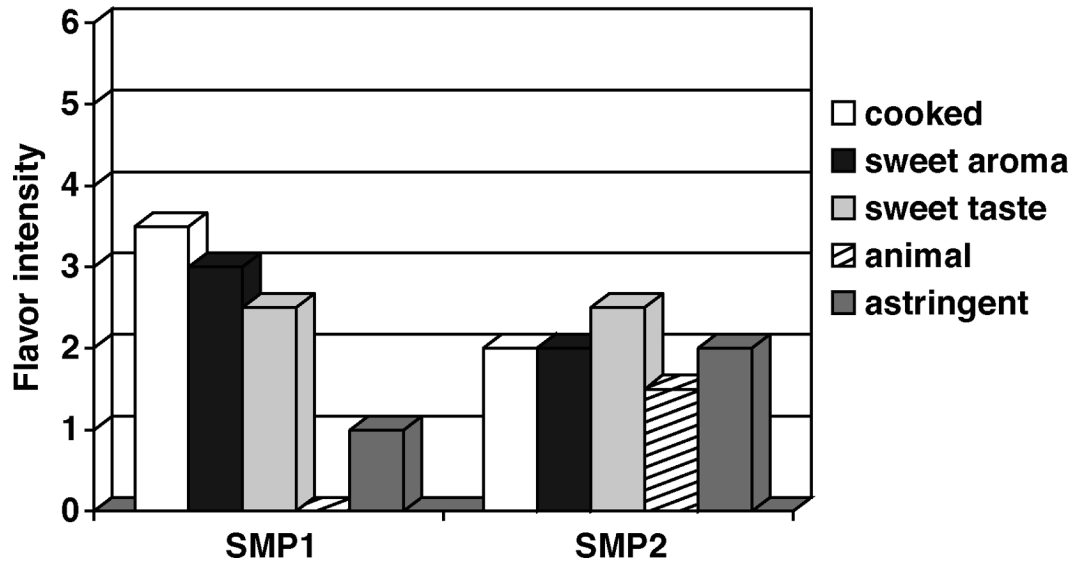


Figure 1. Descriptive sensory profiles provided by a trained descriptive panel for the 2 skim milk powders (SMP) within 1 mo of production. Both products received the same grade (US Extra) from a licensed USDA grader. Flavor intensity was scored on a 15-point universal intensity scale. Most dairy flavors fall between 0 and 4 (Drake, 2004).

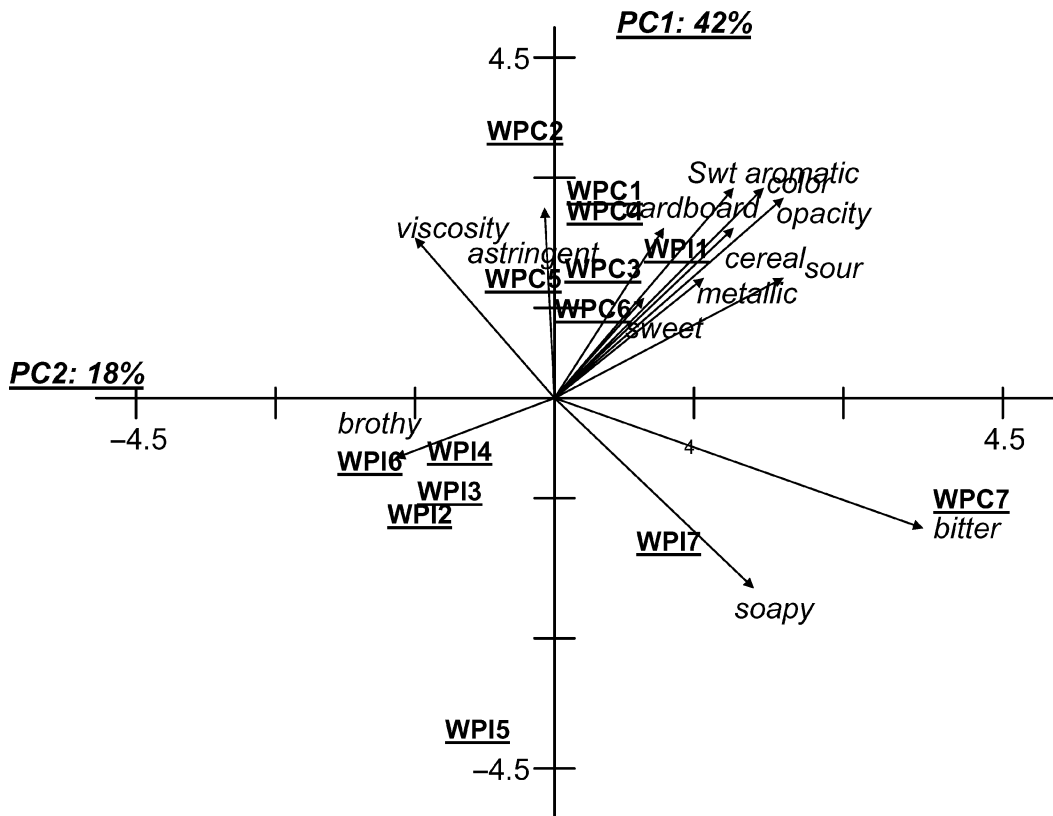


Figure 2. Principal component (PC) biplot of descriptive analysis of whey proteins (PC1 and PC2). Percentage following PC explains amount of variability depicted by each PC on each axis. WPC = whey protein concentrate; WPI = whey protein isolate. Taken from Russell et al. (2006). Reproduced with permission from Wiley Blackwell.

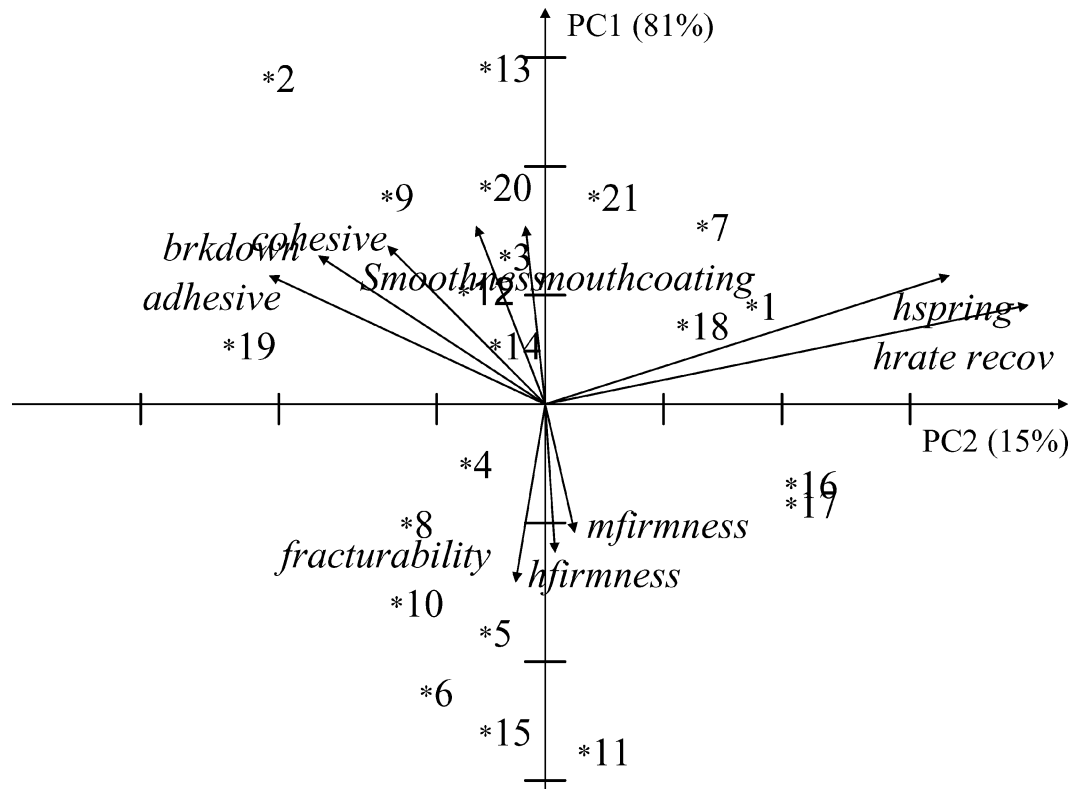


Figure 3. Principal component (PC) plot of descriptive texture attributes of Gouda cheeses. Numbers represent cheeses. Cheeses 1, 16, 17, and 18 are cheeses with reduced-fat contents. Taken from Yates and Drake (2007). Reproduced with permission from Wiley Blackwell.

between sensory and instrumental measurements, and enhance understanding of the consumer. Clear definitions and references for trained descriptive panel attributes facilitate comparison with other studies and instrumental analyses and provide a platform that can be further expanded and applied. The sensory instrument then becomes applicable to a wide array of applications. Drake et al. (2001) developed a sensory language for cheese flavor. The language was developed specifically for Cheddar cheese but once the base language was identified, it was subsequently applied to other cheeses including Swiss, Mozzarella, Parmesan, and goat's milk cheeses with minor modifications (Park et al., 2006; Liggett et al., 2007; Table 2). Drake et al. (2002) demonstrated that a defined cheese flavor language could be used by panels at multiple locations to provide identical results for the same samples. This same defined language was used for comparison and calibration with other descriptive panels (Drake et al., 2005).

A trained descriptive panel plays a critical role in flavor chemistry research by interpretation of volatile compound analysis results. Many volatile components are simply volatile organic compounds and do not play crucial roles in flavor (Drake et al., 2006). Sensory analysis with a trained sensory panel allows the researcher

to interpret instrumental results and pinpoint which volatile or nonvolatile compounds are crucial for specific flavors or basic tastes (Suriyaphan et al., 2001; Avsar et al., 2004; Drake, 2004; Carunchia Whetstine et al., 2005, 2006a,b; Drake et al., 2006; Wright et al., 2006; Carunchia Whetstine and Drake, 2007). Without sensory analysis, there is no relation to flavor, and volatile analysis is simply a list of volatile organic compounds present in the sample. Similar work with sensory analysis can be used to interpret instrumental measurements of physical properties and determine exactly how they relate to sensory perception of texture (Foegeding and Drake, 2007).

One final use of descriptive analysis, other than enhanced product understanding and identification of relationships to instrumental analyses, is to understand consumer perception. In many cases, the reasons consumers like or prefer a product is not clear unless descriptive analysis is applied to the same set of products. By using descriptive analysis, we know the specific sensory or texture profiles of the product; with consumer tests, we know which products consumers like or prefer. For a small number of products or treatments, we can closely examine sensory profiles of well-liked products and can infer why they are liked. Often, the goal is

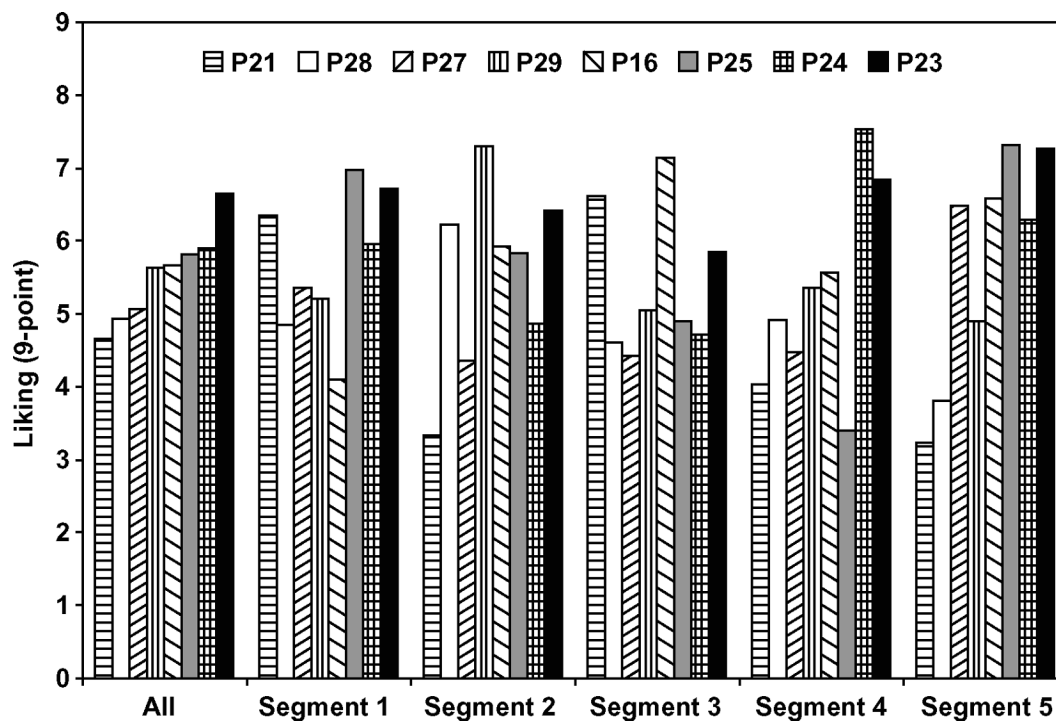


Figure 4. Overall acceptability scores for 6 butters (P16, P21, P23, P24, P25, P27) and 2 spreads (P28, P29) within different identified consumer segments. Taken from Krause et al. (2007).

larger than simply understanding why a specific product is preferred over a few others. Instead, identification of drivers of consumer liking is desired. For this specific goal, a wide range of a particular product is profiled by a trained sensory panel. Selected products are then presented to consumers to obtain liking information. The 2 sets of data are combined in a multivariate statistical technique called preference mapping. A minimum of 8 products with variable trained panel profiles is generally recommended to obtain a robust statistical model (Lawless and Heymann, 1999e). For example, if all products were liked, it would not be possible to identify drivers of liking. This approach has been applied to identify specific consumer likes and dislikes with many dairy products (Jack et al., 1993; Hough and Sánchez, 1998; Lawlor and Delahunty, 2000; Murray and Delahunty, 2000; Richardson-Harman et al., 2000; Thompson et al., 2004; Young et al., 2004; Krause et al., 2007).

The power of these preference-mapping studies is that specific consumer groups with specific likes and dislikes are identified. Figure 4 demonstrates the application of this technique to butter (Krause et al., 2007). Five consumer segments were identified, each with very distinct likes and dislikes for specific butter flavors. The first grouping on the figure shows the mean liking profiles when all product scores were averaged across

all 160 consumers. Note that P23 was the best-liked butter across all consumers. However, P23 was not the best-liked product within each of the 5 consumer clusters, most notably segments 2, 3, and 4. Such information would be overlooked without application of this technique. Specific consumer likes and dislikes can be further linked to demographic criteria (age, income, education, etc.) and product usage habits so that the specific desires and needs of a particular consumer population or segment can be distinguished for effective marketing or product development.

Common Misuses/Abuses of Sensory Tools

Sensory analysis is often written off by individuals and companies as a subjective or haphazard tool due to unsatisfactory results. Generally, the source of the dissatisfaction is based in ignorance of the availability of appropriate sensory analysis tools, which results in selection of the wrong tool or test or misuse of the right tool. The following list represents common mistakes and misuses of sensory analysis.

1. Trained panel results not replicated. The trained sensory panel is an analytical instrument. As with any instrument, results are replicated to ensure reproducibility. This means that each panelist must evaluate each sample at least twice. In some

cases, more replications are conducted by each panelist. More replications are generally conducted when there are very small differences between samples, when the panel is not highly trained, or when the nature of the modality or parameter measured is variable or fatiguing.

2. The sensory language used is not defined or referenced. A well-defined sensory language is crucial not only to have a sensitive and reproducible panel, but also for interpretation or replication of the results or for establishing relationships with consumer liking or instrumental measurements. One word can mean different things to different people. It is crucial to define lexicon terms (Drake and Civille, 2003; Drake et al., 2005; Foegeding and Drake, 2007).
3. Dairy judges are used in place of analytical or affective sensory testing. Dairy judging was designed to provide a rapid measurement of general quality based on predetermined (and common at the time) defects. It was not designed to be an analytical research tool. The defects and their intensities are not well defined and are thus quite variable and difficult to reproduce. Point deductions from quality are not equivalent to different defects, and parametric statistics are not possible. Defects are not necessarily objectionable to the consumer and the quality score generated does not necessarily have any relevance to consumer acceptance. Only predetermined defects are scored, not the entire flavor or texture profile of the product, as would be the case with descriptive analysis. The net result is that very little information relevant to a research or marketing project objective is obtained. The solution is to select a sensory tool appropriate for the research or marketing objective. Several appropriate tests are available and many are simple and easy to conduct.
4. Trained panel is used to measure liking, acceptance, or preference. Once panelists are trained to identify and quantify attributes in products (or grades and defects such as with product judging), they are no longer typical consumers. As such, what they like or prefer generally is no longer relevant or comparable to those of the average consumer. A further issue addressed below is that at least 50 consumers are needed to determine preference or acceptance of a product with any degree of certainty.
5. Low numbers of individuals are used for consumer testing. When preference or acceptance (liking) tests are conducted, the objective is to determine or infer consumer liking or preference. Consumers

are, by their nature, highly variable based on product usage, geographical location, ethnicity, nationality, and so on. As such, a large number of measurements is necessary to have any degree of confidence in the results. Testing with 100 to 500 consumers at multiple locations is conducted by large companies seeking to make sound market decisions. Obviously, research studies do not always have the luxury, due to finances or sample availability, of testing large numbers of consumers. However, it is widely agreed that a minimum of 50 consumers of the product should be polled to make sound conclusions regarding liking or preference (Lawless and Heymann, 1999e).

6. A difference test is used to imply degree of difference, type of difference, or preference. As stated previously, the only thing a difference test can be used for is to determine if a difference exists between 2 or more products. For any additional information, an additional sensory test must be conducted.
7. "Industry does it, so it must be ok." This statement is widely used as a fail-safe excuse for any number of mistakes. This is an invalid issue for several reasons ranging from the old "if-your-friends-jump-off-the-bridge,-you-should-follow-them" adage to the issue that not all companies conduct sound product development and research or even sound sensory analysis. The largest and most successful food (and nonfood) companies have large sensory and market research divisions or make use of a large number of sensory consulting firms. Clearly, much attention is given to appropriate selection of sensory tests and appropriate use of the selected tests.
8. Poor planning. As with any experiment, an appropriate plan, timeline, experimental design, and statistical analysis are crucial for success. Appropriate planning ensures time for proper training of panelists and selection of the appropriate sensory test.

SUMMARY

Sensory analysis is an invaluable set of methods for research and marketing. Knowledge of product variability, stability, comparison to competitor product(s), relationships to instrumental analyses and consumer understanding are all requirements for a successful product. Sensory analysis techniques alone can provide the answers to all of these questions. An array of sensory tools exists, and often a particular goal or research objective may be achieved by more than one correctly conducted sensory test. Many of these tests are simple

and easy to apply, whereas others are more complex and require training and experience. Selection of the appropriate sensory test and testing conditions provides powerful results.

ACKNOWLEDGMENTS

The financial support of the California Dairy Research Foundation and Dairy Management, Inc. are gratefully acknowledged. Paper FSR 07-18 of the journal series of the Department of Food Science, North Carolina State University (Raleigh). The use of trade names does not imply endorsement nor lack of endorsement of those not mentioned.

REFERENCES

- ASTM (American Society for Testing and Materials). 1992. Standard practice for determination of odor and taste thresholds by a forced-choice method of limits. E-679-91. Pages 35–39 in Annual Book of Standards. 15.07. ASTM, Philadelphia, PA.
- Avsar, Y. K., Y. Karagul-Yuceer, M. A. Drake, T. K. Singh, Y. Yoon, and K. R. Cadwallader. 2004. Characterization of nutty flavor in Cheddar cheese. *J. Dairy Sci.* 87:1999–2010.
- Bodyfelt, F. W., M. A. Drake, and S. A. Rankin. Developments in dairy foods sensory science and education—From student contests to impact on product quality. *Int. Dairy J.* (accepted)
- Bodyfelt, F. W., J. Tobias, and G. M. Trout. 1988. The Sensory Evaluation of Dairy Products. Van Nostrand Reinhold, New York, NY.
- Brown, J. A., E. A. Foegeding, C. R. Daubert, and M. A. Drake. 2003. Changes in rheological and sensorial properties of young cheeses as related to maturation. *J. Dairy Sci.* 86:3054–3067.
- Carunchia Whetstine, M. E., K. R. Cadwallader, and M. A. Drake. 2005. Characterization of aroma compounds responsible for the rosy/floral flavor in Cheddar cheese. *J. Agric. Food Chem.* 53:3126–3132.
- Carunchia Whetstine, M. E., and M. A. Drake. 2007. The flavor and flavor stability of skim and whole milk powders. Pages 217–252 in *Flavor Chemistry of Dairy Products*. K. R. Cadwallader, M. A. Drake, R. McGorin, ed. ACS Publishing, Washington, DC.
- Carunchia Whetstine, M. E., M. A. Drake, J. B. Broadbent, and D. J. McMahon. 2006b. Enhanced nutty flavor formation in Cheddar cheese made with a “malty” *Lactococcus lactis* adjunct culture. *J. Dairy Sci.* 89:3277–3284.
- Carunchia Whetstine, M. E., M. A. Drake, B. K. Nelson, and D. Barbano. 2006a. Flavor profiles of full fat, reduced fat and cheese fat made from aged Cheddar with the fat removed using a novel process. *J. Dairy Sci.* 89:505–517.
- Carunchia Whetstine, M. E., P. J. Luck, M. A. Drake, E. A. Foegeding, P. D. Gerard, and D. M. Barbano. 2007. Characterization of flavor and texture development within large (291-kg) blocks of milled and stirred curd Cheddar cheese. *J. Dairy Sci.* 90:3091–3109.
- Chambers, D. H., A. A. Allison, and D. Chambers. 2004. Training effects on performance of descriptive panelists. *J. Sens. Stud.* 19:486–489.
- Childs, J. L., M. D. Yates, and M. A. Drake. 2007. Sensory properties of meal replacement bars and beverages made from whey or soy proteins. *J. Food Sci.* 72:5425–5434.
- Childs, J. L., J. L. Thompson, S. L. Lillard, T. K. Berry, and M. A. Drake. Consumer perception of whey and soy protein in meal replacement products. *J. Sens. Stud.* (accepted)
- Cotunga, N., and C. E. Vickery. 2004. Children rate the summer food service program. *Family Economics Nutr.* 16:3–12.
- Delahunty, C. M., and M. A. Drake. 2004. Sensory character of cheese and its evaluation. Pages 455–487 in *Cheese: Chemistry, Physics and Microbiology*. Vol. 1. General Aspects. 3rd ed. P. F. Fox, P. L. H. McSweeney, T. M. Cogan, and T. P. Guinee, ed. Elsevier, London, UK.
- Drake, M. A. 2004. Defining dairy flavors. *J. Dairy Sci.* 87:777–784.
- Drake, M. A., and G. V. Cville. 2003. Flavor Lexicons. *Comp. Rev. Food Sci.* 2:33–40.
- Drake, M. A., P. D. Gerard, S. Wright, K. R. Cadwallader, and G. V. Cville. 2002. Cross validation of a sensory language for Cheddar cheese. *J. Sens. Stud.* 17:215–229.
- Drake, M. A., Y. Karagul-Yuceer, K. R. Cadwallader, G. V. Cville, and P. S. Tong. 2003. Determination of the sensory attributes of dried milk powders and dairy ingredients. *J. Sens. Stud.* 18:199–216.
- Drake, M. A., S. McIngvale, P. D. Gerard, K. R. Cadwallader, and G. V. Cville. 2001. Development of a descriptive language for Cheddar cheese. *J. Food Sci.* 66:1422–1427.
- Drake, M. A., R. E. Miracle, A. D. Caudle, and K. R. Cadwallader. 2006. Relating Sensory and Instrumental Analyses. Pages 23–55 in *Sensory-Directed Flavor Analysis*. R. Marsili, ed. CRC Press, Taylor and Francis Publishing, Boca Raton, FL.
- Drake, M. A., M. D. Yates, P. D. Gerard, C. M. Delahunty, E. M. Sheehan, R. P. Turnbull, and T. M. Dodds. 2005. Comparison of differences between lexicons for descriptive analysis of Cheddar cheese flavour in Ireland, New Zealand, and the United States of America. *Int. Dairy J.* 15:473–483.
- Foegeding, E. A., and M. A. Drake. 2007. Sensory and mechanical properties of cheese texture. *J. Dairy Sci.* 90:1611–1624.
- Gacula, M., and S. Rutenbeck. 2006. Sample size in consumer test and descriptive analysis. *J. Sens. Stud.* 29:121–145.
- Greene, J. L., K. Bratka, M. A. Drake, and T. H. Sanders. 2006. Effectiveness of category and line scales to characterize consumer perception of fruity fermented flavor in peanuts. *J. Sens. Stud.* 21:146–154.
- Herr, P. M., and C. M. Page. 2004. Asymmetric association of liking and disliking judgments: So what’s not to like? *J. Consum. Res.* 30:588–601.
- Hough, G., and R. Sánchez. 1998. Descriptive analysis and external preference mapping of powdered chocolate milk. *Food Qual. Pref.* 9:197–204.
- Hough, G., I. Wakeling, A. Mucci, E. Chambers, I. Mendez Gallardo, and L. Rangel Alves. 2006. Number of consumers necessary for sensory acceptability tests. *Food Qual. Pref.* 17:522–526.
- IFT/SED (Int. Food Technol./Sens. Eval. Div.). 1981. Sensory evaluation guide for testing food and beverage products. *Food Technol.* 35:550–559.
- Jack, F. R., A. Paterson, and J. R. Piggot. 1993. Relationships between rheology and composition of Cheddar cheeses and texture as perceived by consumers. *Int. J. Food Sci. Technol.* 28:293–302.
- Jiamyanguyen, S., J. F. Delwiche, and W. J. Harper. 2002. The impact of wood ice cream sticks’ origin on the aroma of exposed ice cream mixes. *J. Dairy Sci.* 85:355–359.
- Jones, V. S., M. A. Drake, R. Harding, and B. Kuhn-Sherlock. Consumer perception of soy and dairy products: A cross-cultural study. *J. Sens. Stud.* (accepted)
- Karagul-Yuceer, M. A. Drake, and K. R. Cadwallader. 2004. Evaluation of the character impact odorants in skim milk powder by sensory studies on model mixtures. *J. Sens. Stud.* 19:1–14.
- Kosa, K. M., S. C. Cates, R. C. Post, and J. Canavan. 2004. Consumers’ attitudes toward labeling food products with possible allergens. *Food Prot. Trends* 24:605–611.
- Krause, A. J., K. Lopetcharat, and M. A. Drake. 2007. Identification of the characteristics that drive consumer liking of butter. *J. Dairy Sci.* 90:2091–2102.
- Kreuger, R. A., and M. A. Casey. 2000. *Focus Groups: A Practical Guide for Applied Research*. 3rd ed. Sage Publications, Thousand Oaks, CA.
- Kroll, B. J. 1990. Evaluating rating scales for sensory testing with children. *Food Technol.* 44:78–86.
- Lawless, H. T., and H. Heymann. 1999a. Physiological and psychological foundations of sensory function. Pages 28–74 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.

- Lawless, H. T., and H. Heymann. 1999b. Discrimination testing. Pages 116–138 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.
- Lawless, H. T., and H. Heymann. 1999c. Measurement of sensory thresholds. Pages 173–204 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.
- Lawless, H. T., and H. Heymann. 1999d. Descriptive analysis. Pages 341–372 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.
- Lawless, H. T., and H. Heymann. 1999e. Consumer field tests and questionnaire design. Pages 480–514 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.
- Lawless, H. T., and H. Heymann. 1999f. Qualitative consumer research methods. Pages 519–547 in *Sensory Evaluation of Food*. 1st ed. Chapman and Hall, New York, NY.
- Lawlor, J. B., and C. M. Delahunty. 2000. The sensory profile and consumer preference for ten specialty cheeses. *Int. Dairy Technol. J.* 53:28–36.
- Liggett, R., M. A. Drake, and J. F. Delwiche. Impact of sensory characteristics on consumer liking of Swiss cheese. *J. Dairy Sci.* (accepted)
- Meilgaard, M. M., G. V. Civille, and T. Carr. 2007a. Overall difference tests: Does a sensory difference exist between samples? Pages 63–104 in *Sensory Evaluation Techniques*. 4th ed. CRC Press, New York, NY.
- Meilgaard, M. M., G. V. Civille, and T. Carr. 2007b. Attribute difference tests: How does attribute x differ between samples? Pages 105–128 in *Sensory Evaluation Techniques*. 4th ed. CRC Press, New York, NY.
- Meilgaard, M. M., G. V. Civille, and T. Carr. 2007c. Determining threshold. Pages 129–139 in *Sensory Evaluation Techniques*. 4th ed. CRC Press, New York, NY.
- Meilgaard, M. M., G. V. Civille, and T. Carr. 2007d. Affective tests: Consumer tests and in-house panel acceptance tests. Pages 255–309 in *Sensory Evaluation Techniques*. 4th ed. CRC Press, New York, NY.
- Moskowitz, H. R., J. H. Beckley, and A. V. A. Resurreccion. 2007. *Sensory and Consumer Research in Food Product Development*. Blackwell Publishing and IFT Press, Ames, IA.
- Murray, J. M., and C. M. Delahunty. 2000. Mapping consumer preference for the sensory and packaging attributes for Cheddar cheese. *Food Qual. Pref.* 11:419–435.
- Murray, J. M., C. M. Delahunty, and I. Baxter. 2001. Descriptive sensory analysis: A review. *Food Res. Int.* 34:461–471.
- Orme, B. K. 2006. *Getting Started with Conjoint Analysis. Strategies for Product Design and Pricing Research*. Research Publishers LLC, Madison, WI.
- Park, Y., P. D. Gerard, and M. A. Drake. 2006. Impact of frozen storage on flavor of caprine milk cheeses. *J. Sens. Stud.* 21:654–663.
- Peryam, D. R., and F. J. Pilgrim. 1957. Hedonic scale method of measuring food preferences. *Food Technol.* 11:9–14.
- Popper, R., W. Rosenstock, M. Schraidt, and B. J. Kroll. 2004. The effect of attribute ratings on overall liking ratings. *Food Qual. Pref.* 15:853–858.
- Rehman, S. U., N. Farkye, and M. A. Drake. Differences between Cheddar cheese manufactured by the milled curd and stirred curd methods using different commercial starters. *J. Dairy Sci.* (accepted)
- Resurreccion, A. V. A. 1998. The consumer panel. Pages 71–91 in *Consumer Testing for Product Development*. Aspen Publishers, Gaithersburg, MD.
- Richardson-Harman, N. J., R. Stevens, S. Walker, J. Gamble, M. Miller, M. Wong, and A. McPherson. 2000. Mapping consumer perceptions of creaminess and liking for liquid dairy products. *Food Qual. Pref.* 11:239–246.
- Robinson, K. M., B. P. Klein, and S. Y. Lee. 2004. Utilizing the R-index measure for threshold testing in model soy isoflavone solutions. *J. Food Sci.* 69:SNQ1–SNQ4.
- Robinson, K. M., B. P. Klein, and S. Y. Lee. 2005. Utilizing the R-index measure for threshold testing in model caffeine solutions. *Food Qual. Pref.* 16:284–289.
- Russell, T., M. A. Drake, and P. D. Gerard. 2006. Sensory properties of whey and soy proteins. *J. Food Sci.* 71:S447–S455.
- Rychlik, M., P. Schieberle, and W. Grosch. 1998. Compilation of thresholds, odor qualities, and retention indices of key food odorants. Deutsche Forschungsanstalt für Lebensmittelchemie and Institut für Lebensmittelchemie der Technischen Universität München, Garching, Germany.
- Schlich, P. 1995. Preference mapping: Relating consumer preferences to sensory or instrumental measurements. Pages 231–245 in *Bi-Ofavour '95. Analysis/Precursor Studies/Biotechnology*. P. Etivant and P. Schreier, ed. INRA Editions, Versailles, France.
- Schutz, H. G., and A. V. Cardello. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *J. Sens. Stud.* 16:117–159.
- Sherlock, M., and T. P. Labuzza. 1992. Consumer perceptions of consumer time-temperature indicators for use on refrigerated dairy foods. *J. Dairy Sci.* 75:3167–3176.
- Singh, T., M. A. Drake, and K. R. Cadwallader. 2003. Flavor of Cheddar cheese: A chemical and sensory perspective. *Comp. Rev. Food Sci.* 2:139–162.
- Suriyaphan, O., M. A. Drake, X. Q. Chen, and K. R. Cadwallader. 2001. Characteristic aroma components of British Farmhouse Cheddar cheese. *J. Agric. Food Chem.* 49:1382–1387.
- Thompson, J. L., M. A. Drake, K. Lopetcharat, and M. D. Yates. 2004. Preference mapping of commercial chocolate milks. *J. Food Sci.* 69:S406–S413.
- van Gemert, L. J. 2003. *Compilations of odour threshold values in air, water, and other media*. Oliemans Punter and Partners BV, Rotterdam, the Netherlands.
- Wright, J. M., M. E. Carunchia Whetstine, R. E. Miracle, and M. A. Drake. 2006. Characterization of a cabbage off-flavor in whey protein isolate. *J. Food Sci.* 71:C91–C96.
- Yates, M. D., and M. A. Drake. 2007. Texture properties of Gouda cheese. *J. Sens. Stud.* (accepted)
- Young, N., M. A. Drake, K. Lopetcharat, and M. McDaniel. 2004. Preference mapping of Cheddar cheeses. *J. Dairy Sci.* 87:11–19.