
Derivation and Evaluation of a Labeled Hedonic Scale

Juyun Lim¹, Alison Wood¹ and Barry G. Green^{2,3}

¹Department of Food Science and Technology, Oregon State University, Corvallis, OR 97331, USA, ²The John B. Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519, USA and ³Department of Surgery (Otolaryngology), Yale Medical School, New Haven, CT 06519, USA

Correspondence to be sent to: Juyun Lim, Department of Food Science and Technology, Oregon State University, 100 Wiegand Hall, Corvallis, OR 97331, USA. e-mail: juyun.lim@oregonstate.edu

Abstract

The objective of this study was to develop a semantically *labeled hedonic scale* (LHS) that would yield ratio-level data on the magnitude of liking/disliking of sensation equivalent to that produced by magnitude estimation (ME). The LHS was constructed by having 49 subjects who were trained in ME rate the semantic magnitudes of 10 common hedonic descriptors within a broad context of imagined hedonic experiences that included tastes and flavors. The resulting bipolar scale is statistically symmetrical around neutral and has a unique semantic structure. The LHS was evaluated quantitatively by comparing it with ME and the 9-point hedonic scale. The LHS yielded nearly identical ratings to those obtained using ME, which implies that its semantic labels are valid and that it produces ratio-level data equivalent to ME. Analyses of variance conducted on the hedonic ratings from the LHS and the 9-point scale gave similar results, but the LHS showed much greater resistance to ceiling effects and yielded normally distributed data, whereas the 9-point scale did not. These results indicate that the LHS has significant semantic, quantitative, and statistical advantages over the 9-point hedonic scale.

Key words: category-ratio scale, labeled hedonic scale, magnitude estimation, 9-point hedonic scale

Introduction

The measurement of hedonic responses is fundamental to understanding the relationship of the chemical senses to food preference and selection. Despite this fact, development of hedonic scaling has lagged behind the development of intensity scaling. For more than half a century, the traditional 9-point hedonic scale (Peryam and Girardot 1952; Peryam and Pilgrim 1957), in its various formats (e.g., labels only, labels with numbers), has been widely used to assess the average degree of liking or disliking of foods or consumer products across a large number of subjects. In recognition of the positive aspects of this scale (e.g., ease of use; for review see Lawless and Heymann 1998), it was adopted by many researchers in the chemical senses and became the dominant tool for measuring hedonic perception. However, the scale yields only rudimentary data on hedonic magnitude and cannot provide meaningful comparisons of hedonic perception between individuals and groups (Bartoshuk et al. 2006). First, because of its categorical structure of the scale (i.e., ratings are limited to 9 categories) and further because of its “inequality of scale intervals and the lack of a zero point” (Peryam and Pilgrim 1957, p. 14), the scale can yield only ordinal- or, at best, interval-level data. Second, the scale

is highly vulnerable to ceiling effects (Stevens and Galanter 1957; Schutz and Cardello 2001), because of both its small number of available categories (4 positive and 4 negative) and the general tendency of subjects to avoid using extreme categories (Hollingworth 1910; Moskowitz 1982). Both of these limitations raise questions about the validity and sensitivity of the scale. In addition, from a statistical standpoint, because the data are categorical and discrete without a true zero point, the types of statistical analyses that can be applied with confidence are limited (i.e., nonparametric analyses). Because nonparametric analyses are insensitive compared with parametric analyses, it is thus a common practice for researchers to treat data obtained with the 9-point scale as if the numbers assigned to the categories were points on a continuum (Peryam and Pilgrim 1957). However, as recognized in one of the original publications of the 9-point hedonic scale (Peryam and Pilgrim 1957), some of the assumptions for parametric analyses (e.g., normality, homogeneity of variance) are often violated (Gay and Mead 1992; Villanueva et al. 2000), especially for ratings near the ends of the scale (Peryam et al. 1960). Accordingly, valid statistical inferences cannot be drawn from category scales

unless a large sample size is used to approximate normality (i.e., the Central Limit Theorem). This has been less of a problem for its application in food science, where about hundreds of individuals are often tested to evaluate liking and disliking of food products. It is more of a problem for application of the 9-point scale in basic research, where subject numbers are typically much smaller.

The method of magnitude estimation (ME; Stevens 1957; Stevens and Galanter 1957), which was originally developed to quantify sensory magnitude, has also been used as a tool for hedonic measurement (Engen and McBurney 1964; Moskowitz 1971, 1982), primarily in basic research. The greatest benefit of this ratio-scaling method is that it can theoretically produce ratio-level data (i.e., the highest level of measurement) (Stevens 1957). Unfortunately, ME does not provide semantic information about sensory experience and thus prevents researchers from making comparisons of individual differences. In addition, the difficult nature of the modulus-free ME task, in which subjects must estimate numerical ratios of sensory experience using an unrestricted range of numbers (Moskowitz 1977), means that the quality of the data obtained with ME often depends on the level of experience or training that subjects have with the method. For these reasons, ME has not been widely used to measure hedonic perception, particularly in applied studies that involve consumers.

More recently, a different type of scale has come into use that aims to maintain the chief advantages of the traditional category scale (i.e., ease of use and the availability of semantic information about sensation magnitude) while raising the level of data the scale provides. Category-ratio scales, such as the CR-10 scale (Borg 1982) or the Labeled Magnitude Scale (LMS) (Green et al. 1993, 1996), are continuous line scales on which the location of verbal descriptors is based on their semantic magnitudes as empirically determined via ME (Stevens 1957; Stevens and Galanter 1957). The key features and properties of these scales are 1) because they were derived and validated using ratio scaling (i.e., ME), they can be assumed to yield ratio-level data equivalent to ME; 2) because they are bounded by “no sensation” and “strongest (or maximal) imaginable sensation” on each end, they enable comparison of individual and group differences within the context of the full range of perceived intensities; and 3) because the positions of their semantic labels have been empirically determined, they provide meaningful semantic information about subjective experience.

As noted above, ME is based on the assumption that subjects are able to make numerical judgments in direct proportion to sensory magnitude (Stevens 1953, 1955). Because there is no way to test this assumption in an absolute sense, some researchers have questioned the validity of the method (Attneave 1962; Birnbaum 1980; Anderson 1982). However, strong evidence in support of ME has come from studies in vision, hearing, and touch that found additivity of sensation magnitude for pairs of independent stimuli (Hellman and

Zwislocki 1964; Marks 1979; Zwislocki 1983; Bolanowski 1987). In addition, Marks and Bartoshuk (1979) combined ME with an intensity-matching task and found additivity of perceived taste intensity for equi-intense components of a mixture. Close agreement in the latter study between ME and a direct matching task provides particularly strong support for the assumption that category-ratio scales that produce psychophysical functions equivalent to those produced by ME yield ratio-level data.

The second and third characteristics of category-ratio scales derive from the semantic information that the scales' verbal labels afford. Because adjectives and adverbs possess psychological magnitude (Moskowitz 1977; Borg 1982; Bartoshuk et al. 2004), positioning labels on a scale in accordance with their magnitudes is essential for obtaining meaningful, quantitative information about subjective experience. As Stevens (1958) pointed out with the classical example of large mice versus small elephants, use of an appropriate and common frame of reference is necessary for making valid comparisons of the perception of different stimuli and for comparing sensory experience across individuals and groups. The positions of the labels relative to one another, and critically, to the end points of the scale, establishes a structured and common frame of reference within which all subjects make their responses and thus within which their experiences can theoretically be meaningfully compared.

Following this logic, Schutz and Cardello (2001) were the first to develop an affective category-ratio scale, which they called the Labeled Affective Magnitude (LAM) scale. The authors extensively reviewed the theoretical foundations of category-ratio scales and patterned the development of the LAM scale after the procedure originally used to create the LMS. However, details of the psychophysical procedure Schutz and Cardello used differed somewhat from those used by Green et al. (1993). First, very few of the subjects who participated in the semantic scaling task that generated the LAM scale (and in the subsequent experiments comparing it to ME and to the 9-point scale) had training or experience with ME. Instead, over 90% of the subjects had experience using traditional, equally spaced category scales in sensory tests of foods. Although it is difficult to know which aspects of the procedure are most critical for scale development, subjects' prior training in the scaling task, and hence the ability to use it properly without confusion or bias, is of obvious importance. Indeed, because it requires the unusual task of assigning numbers to express the ratio of perceptual experiences to one another, it has been shown that practice with ME is necessary to obtain reliable ratios (Moskowitz 1977). Secondly, subjects were instructed to rate the hedonic magnitudes of various verbal phrases as they are commonly used to describe the degree of liking of foods rather than rating them in the broader context of all hedonic experiences. Although the food context provided an appropriate frame of reference to make comparisons among food samples, it is unclear that it provided a valid context in which

to compare the hedonic value of nonfood items or to compare individual and group differences in hedonic perception.

Another recently published hedonic scale whose development was also patterned after the LMS is the “Oral Pleasantness and Unpleasantness Scale” (OPUS) (Guest et al. 2007). Unlike the LAM scale, the OPUS (comprising 2 separate scales of pleasantness and unpleasantness) was developed using subjects who were trained and given experience in ME prior to the semantic scaling task. However, similar to the LAM scale, the perceptual context created for the semantic scaling task was relatively narrow, limited in this case to oral sensations, with painful sensations purposely avoided. In addition, to accommodate the development of scales of wetness and dryness at the same time, the semantic context contained mainly examples of perceived intensity. Thus, subjects rated semantic descriptors for wetness, dryness, pleasantness, and unpleasantness amidst examples of nonpainful oral sensations whose intensities they also rated. Interestingly, the semantic structure of the OPUS is similar to the LAM scale (Guest et al. 2007).

A third category-ratio scale that is being used for hedonic scaling in the chemical senses is the general version of the LMS (gLMS) modified to be a bipolar scale (Bartoshuk et al. 2004). Based on the assumption that hedonic magnitude and perceived intensity have a similar scalar structure (Moskowitz and Chandler 1977; Bartoshuk et al. 2004), the bipolar gLMS has “neutral” at its midpoint with negative descriptors to the left bounded by “strongest imaginable displeasure of any kind,” and positive descriptors to the right bounded by “strongest imaginable pleasure of any kind” (Duffy et al. 1999). Because the original gLMS was intended to measure sensory intensities, with the exception of “moderate,” the semantic labels of the bipolar gLMS have no direct counterparts on the LAM scale or OPUS. However, when the structure of the bipolar gLMS is compared with the structures of the other 2 hedonic category-ratio scales, “moderate” on the bipolar gLMS is located much closer to neutral than is “moderately” on the other 2 scales. This disparity may mean that the underlying hedonic and intensity continua have different semantic structures or, alternatively, that the difference between the scales is a byproduct of the different psychophysical procedures that were used to generate them.

The main objective of the present study was, therefore, to use the original LMS psychophysical procedure to develop a potentially new hedonic category-ratio scale. Such a scale would, theoretically, have the same quantitative and semantic advantages of the gLMS but be optimized for measuring hedonic magnitude. In addition, the semantic structure of the resulting scale would help to determine whether the underlying hedonic and intensity continua are similar or different. After finding the new scale had a semantic structure that differed from the LAM scale and OPUS, we went on to test the scale 1) against ME to evaluate the validity of its hedonic descriptors and thus its potential ratio properties, and 2)

against the traditional 9-point category scale to assess its resistance to ceiling effects and its relative statistical power.

Experiment 1: scale derivation

The objective of the first experiment was to create a scale by quantifying the semantic values of terms commonly used to describe liking and disliking of sensations of all kinds within the full range of experienced and imaginable hedonic magnitudes using ME.

Materials and methods

Subjects

A total of 54 subjects (35 females and 19 males) between 18 and 40 years of age (mean = 24 years old) were recruited on the Yale University Campus and were paid to participate. The experimental protocol was approved by the Yale University human investigations committee, and subjects gave written informed consent. All participants were non-smoking native English speakers who reported that they were free from deficits in taste or smell. Subjects were asked to refrain from eating or drinking for at least 1 h prior to their scheduled session when tasting was involved. Some of the subjects had prior experience with the general version of the LMS (Green et al. 1993; Bartoshuk et al. 2002) and with rating the intensity of oral and/or temperature sensations in a laboratory context. However, none were familiar with either ME or the 9-point hedonic scale.

Practice stimuli

A variety of different taste stimuli (see Table 1) were used to train the subjects on how to use ME during the practice session (see below). The chemical stimuli were prepared weekly from reagent grade compounds using deionized water. All the stimuli were stored at 4–6 °C prior to use and were served at room temperature (20–22 °C).

Procedure

Each subject attended 2 sessions on separate days. The purpose of the first session was to learn the method of ME and to practice using it to rate sensations. The purpose of the second session was to use ME in a semantic scaling task to quantify the hedonic magnitude of several different descriptors of liking and disliking that would be used to construct the scale.

Magnitude estimation practice procedure. The first session began by instructing subjects in modulus-free ME. The subjects were then asked to assign numbers to hand spans generated by the experimenter. This served to indicate to the experimenter whether the subject understood the task and also gave subjects practice assigning a variety of numbers to a wide range of “magnitudes.” Subjects were then asked

Table 1 Taste stimuli used to train subjects on how to use ME during the practice session

	Stimuli	Source	Details
Practice I (intensity ratings)	Sucrose	J.T. Baker	0.032, 0.056, 0.18, 0.56, 1.0 M
	Sodium chloride	J.T. Baker	0.032, 0.056, 0.18, 0.56, 1.0 M
Practice II (hedonic ratings)	Sucrose	J.T. Baker	0.056, 1.0 M
	Sodium chloride	J.T. Baker	0.056, 1.0 M
	Citric acid	Pfaltz & Bauer Inc.	17, 56 mM
	Quinine chloride	Fisher Scientific	0.1, 1.0 mM
	Monosodium glutamate	Sigma–Aldrich	80, 180 mM
Practice III (hedonic ratings)	Salted cracker	Nabisco, Inc.	Unsalted tops Premium saltine crackers
	M & M's chocolate	Mars, Inc.	Brown color
	Jelly beans	Jelly Belly Candy Company	Cherry and popcorn flavor
	Starburst	Mars, Inc.	Orange flavor

to rate the taste intensity of 5 concentrations each of sucrose and NaCl solutions (Table 1) using ME. The taste stimuli were presented twice each in a pseudorandom sequence (Practice I). Before each taste stimulus was presented, subjects rinsed at least twice with deionized water (37 ± 0.5 °C). The stimulus was presented in 10-ml aliquots and was held in the front of the mouth for 5 s. They then expectorated the stimulus and rated its intensity magnitude. After a 3-min break, a series of 10 taste stimuli (Table 1, Practice II) was presented using the same procedure, but this time, the task was to rate their hedonic magnitudes using ME. For each stimulus, subjects first indicated the hedonic valence (like, dislike, or neutral) of the sensation and then rated its hedonic magnitude. In the last part of the practice session (Practice III), subjects were presented with 5 commercially available food products (see Table 1). They again indicated the hedonic valence of the flavor of each product and then assessed the magnitude of liking or disliking. For the food products, subjects were asked to consume the samples the way they usually do and then rate the magnitude of liking or disliking. Subjects again rinsed at least twice during a 1-min intertrial interval.

Scale development procedure. In the second session, subjects estimated the magnitude of 10 adjectives that described different levels of liking and disliking within a context of the full range of imagined hedonic magnitudes. Based on published data (Jones et al. 1955; Jones and Thurstone 1955; Schutz and Cardello 2001), we selected the adjectives “slightly,” “moderately,” “very much,” and “extremely” as unambiguous (i.e., unlikely to be confused) scale descriptors of both positive and negative hedonic tone. In addition, the phrases “most liked sensation imaginable” and “most disliked sen-

sation imaginable” were included to establish equivalent end points that encompass the full range of possible hedonic experiences. To provide a perceptual context in which to interpret the given phrases and thus to rate the magnitude of semantic descriptors of hedonic value, a list of 40 examples of familiar liked and/or disliked sensations (see Table 2) was developed. Subjects were given instructions to rate the degree of liking or disliking of each of the imagined sensations by entering numerical magnitude estimates on the keyboard after each item was read to them by the experimenter. Subjects then rated the full list ($N = 40$, 20 for each liked and disliked sensation) once again with the verbal descriptors ($N = 10$) presented in random order after every fourth example sensation. Thus, the subjects had made 22 magnitude estimates of generally liked or disliked sensations (the full list plus the first 2 repeated examples for each liked and disliked example) before they were exposed to the first descriptor and then continued to rate the examples a second time as they appeared between descriptors. To avoid causing confusion by asking subjects to make positive and negative magnitude estimates, the examples were grouped according to whether they were expected to generally be liked or disliked. Half of the subjects rated generally liked examples first and half rated generally disliked examples first. Subjects were instructed that they may sometimes find they dislike a generally liked example (or vice versa) and that if so they should indicate this by entering “x” rather than making a numerical rating. All testing was conducted in a psychophysics laboratory on a one-to-one basis.

Data analysis

Paired *t*-tests were carried out to examine if the positive and negative dimensions for each descriptor were statistically different by using Statistica 8.0 (StatSoft Inc.).

Table 2 Imagined sensations used for scale development

Liked sensations	Disliked sensations
The taste of water	Seeing your least favorite color
The feel of washing your hands in warm water	The feel of a minor scratch
The taste of plain bread	Touching coarse sandpaper
The smell of clean laundry	The taste of room temperature soda
The feel of a comfortable chair	A minor cough
The sound of a beautiful bird song	The taste of a soggy potato chip
Listening to your favorite song	The feel of tight shoes
The feel of a warm bed in winter	The smell of dirty laundry
The taste of an iced drink on a hot day	Stinging eyes from cutting an onion
The feel of a tropical breeze	Sore feet from walking
The smell of a rose	The itch of a mosquito bite
The sound of the ocean	Stubbing your toe
The softness of baby's skin	The smell of bad body odor
The feel of bathwater	A very loud TV or radio
The feel of pure silk	The smell of garbage
The taste of your favorite chocolate	The sound of fingernails dragging across a blackboard
The warmth of a fireplace in winter	A loud ambulance siren
The taste of cold water when you are thirsty and hot	The feel of scalding hot water
The taste of your favorite dessert	The taste of rotten food
The feel of a massage	The smell of vomit

Results

Although most subjects had no difficulty in performing the scale development task, 5 either gave adjacent descriptors inverse ratings or were confused about number usage when rating the negative descriptors (i.e., larger numbers were used for smaller negative hedonic magnitudes). The data from those subjects were discarded and the data from the remaining 49 subjects were used to construct the scale. Visual inspection of a frequency plot suggested that the distributions of responses to each descriptor were approximately log-normal, which is expected for ME data (Stevens 1957; Butler et al. 1987). Thus, log means of the magnitude estimates were calculated across subjects. Table 3 contains the logged mean magnitude estimates with their 95% confidence intervals (CIs) and the geometric means (i.e., the antilogs of the log means) for the 5 positive and 5 negative descriptors. Figure 1 shows the “Labeled Hedonic Scale” (LHS) that was constructed based on the geometric means. Although semantic magnitudes for the most extreme descrip-

tors (i.e., “most liked sensation imaginable” and “most disliked sensation imaginable”) were not exactly the same, they were not significantly different (see Table 3). Indeed, the geometric means for each descriptor for the positive and negative dimensions were very similar and not statistically different (paired *t*-tests, $P > 0.05$), indicating that “liking” and “disliking” are, on average, symmetrical rather than asymmetrical (Moskowitz 1977; Schutz and Cardello 2001). The LHS is therefore a symmetrical, bipolar scale. For purposes of quantifying hedonic ratings, the scale has an arbitrary range from -100 to $+100$ and is anchored at the center (0) by neutral. Because it is intended as a semantic scale and because numbers could distract subjects from making ratings based solely on its semantic labels, no numbers are placed on the LHS.

The semantic labels of the LHS are not evenly spaced, indicating that the hedonic intervals between neighboring categories are unequal. For example, the difference in geometric means between the descriptors “like slightly” and “like moderately” is 11.6 units, whereas the difference between “like very much” and “like extremely” is 21.3 units. However, given the manner in which the LHS was constructed, it is more appropriate to describe the relationship among descriptors in terms of ratios rather than differences. The ratios between any 2 descriptors can be calculated from either their geometric means or their standardized scale values (Table 3). For example, a sensation described as “like extremely” is approximately 10.5 times better liked than a sensation described as “liked slightly,” and a sensation “disliked moderately” is about 2.4 times less disliked than a sensation “disliked very much.”

A comparison of the semantic structure of the LHS with those of the LAM scale (Schutz and Cardello 2001) and OPUS (Guest et al. 2007) revealed that labels on the LHS are much less evenly spaced than on either of the other 2 scales (Figure 2). For example, when the scales are standardized to a range of ± 100 , the label “like moderately” has a scale value of $+17.8$ on the LHS, whereas the same label has a value of $+36.2$ on the LAM scale and $+40.8$ on the OPUS. Values for “dislike moderately” on the other scales show the same marked deviation from the LHS, as do the remaining labels that all 3 scales have in common. Interestingly, the location of “moderate” on the gLMS is remarkably close to “like moderately” on the LHS ($+17.1$ compared with $+17.8$). However, because none of the other semantic labels of the gLMS are directly comparable to labels on the LHS (e.g., “barely detectable”), a comparison between the overall structures of these 2 scales could not be made.

Experiment 2: comparison of scales

After finding that the LHS had a different semantic structure than the other hedonic category-ratio scales, we went on to compare hedonic ratings collected with it to data collected with ME and with the traditional 9-point hedonic scale, which is the most commonly used hedonic scale. Two scale

Table 3 Semantic phrases, the means of logged magnitude estimates ± 95% CIs, and the geometric mean ratings for the 5 positive and 5 negative descriptors

Positive phrases				Negative phrases			
Descriptors	Log mean ± 95% CI	Geometric mean	Scale value	Descriptors	Log mean ± 95% CI	Geometric mean	Scale value
Most liked sensation imaginable	2.20 ± 0.16	156.73	100.00	Most disliked sensation imaginable	2.21 ± 0.17	161.24	-100.00
Like extremely	2.01 ± 0.14	103.00	65.72	Dislike extremely	2.01 ± 0.15	101.41	-62.89
Like very much	1.84 ± 0.14	69.64	44.43	Dislike very much	1.83 ± 0.14	67.05	-41.58
Like moderately	1.45 ± 0.13	27.93	17.82	Dislike moderately	1.45 ± 0.14	28.37	-17.59
Like slightly	0.99 ± 0.12	9.80	6.25	Dislike slightly	0.98 ± 0.14	9.55	-5.92

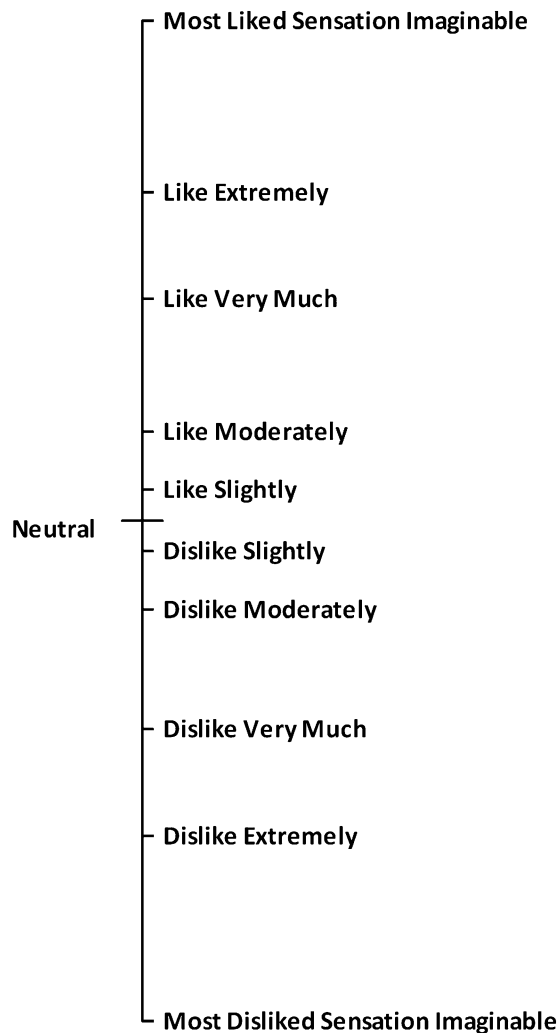


Figure 1 The LHS constructed from the geometric means of magnitude estimates of the 10 semantic descriptors.

properties were of special interest: 1) the potential to yield ratio-level data and 2) the presence or absence of ceiling effects. If the results from the LHS were found to be statisti-

cally comparable with those from ME, the LHS could be assumed to yield ratio-level data equivalent to ME. The 9-point hedonic scale is inherently more vulnerable to ceiling effects because its end points do not accommodate ratings of stimuli that are judged to be more than extremely liked or disliked. In addition, subjects tend to avoid using the end categories on the 9-point hedonic scale (i.e., the “end effect”) (Stevens and Galanter 1957; Moskowitz 1982). To cover a wide range of liked and disliked sensations, a variety of food item names were used as test stimuli.

Materials and methods

Subjects

A total of 55 subjects (44 females and 11 males) between 18 and 50 years of age (mean = 22 years old) were recruited on the Oregon State University campus and were paid to participate. None of the subjects had prior experience with a psychophysical test in a laboratory context and thus none were familiar with any scale used during the experiment. The experimental protocol was approved by the Oregon State University Institutional Review Board, and subjects gave written informed consent. All participants were nonsmoking native English speakers who reported that they were free from deficits in taste or smell.

Stimuli

Twenty-six food item names (see Table 4) were selected to span a much wider range of liked and disliked flavors and foods than would be possible to present in a laboratory setting. This approach enabled us to obtain ratings over the widest possible range of hedonic values for each method under the same stimulus condition.

Procedure

Each subject attended 3 sessions on separate days. During each session, subjects used one of the 3 scale types to rate their liking or disliking for the 26 food item names: The LHS as seen in Figure 1, the standard 9-point hedonic scale

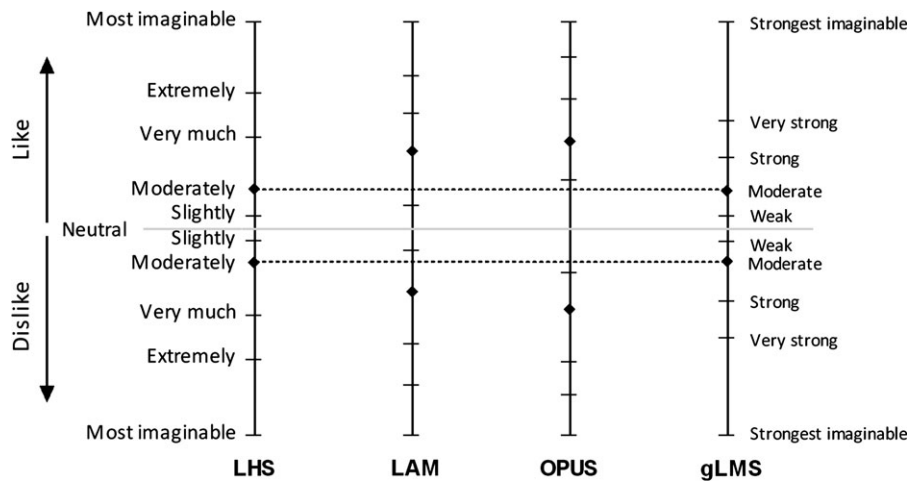


Figure 2 Shown for comparison are the locations of semantic descriptors on the LHS, the LAM, the OPUS, and the bipolar form of the (gLMS). Filled diamonds indicate the location of moderately on each scale, which is the only semantic descriptor other than neutral that is common to all of the scales. Horizontal dotted lines intersect the other scales at the locations of like moderately and dislike moderately on the LHS. Tick marks indicate the locations of the 4 other positive and negative descriptors that are “semantically equivalent” on the LHS, LAM, and OPUS. The remaining 4 descriptors of the gLMS, which have no direct counterparts on the other 3 scales, are shown on the right.

(with labels only) (Peryam and Pilgrim 1957), and the hedonic format of modulus-free ME (Moskowitz 1977, 1982). The order of presentation of the methods was counter-balanced across subjects so that each method appeared equally often in the first, second, and third sessions. At the beginning of each session, verbal instructions were given for the method that was to be tested in that session. To make certain that subjects understood the concept of ratio estimation for the ME task, the instructions included a series of practice ratings to a range of hand spans produced by the experimenter. All ratings were made on paper ballots, with each stimulus rated on a separate ballot. Rating booklets were prepared for each session, and each page of the booklet contained the appropriate scale (or place to write magnitude estimates) and one of the 26 food names to be rated. All testing was conducted in a psychophysics laboratory on a one-to-one basis.

Data analysis

Prior to statistical analysis, the data from the scaling methods were transformed to enable comparisons across methods. To be able to make direct comparisons with the LHS, the category values for the 9-point hedonic scale were transformed into a range from -100 (dislike extremely) to $+100$ (like extremely) in increments of 25 units corresponding to its 4 positive and 4 negative labels. Visual inspection of individual data from ME revealed that, as is typically found, they deviated from normality due to individual differences in numerical ratings. The ME data were therefore normalized across subjects by dividing the grand mean of the absolute values of ratings of all subjects into the mean of the absolute values of the ratings for each subject, then multiplying the

individual subject ratings by the resulting factor for each subject. After the normalization process, with the exception of 5 food item names (A, L, R, T, and Z), the ME data followed the normal distribution. To compare the LHS and ME directly, a standardization procedure (Moskowitz 1977) was further carried out that equalized the grand means of absolute values from the 2 scales. This standardization procedure avoids differences between methods based solely on number usage (Green et al. 1993). It is important to note that it was necessary to use absolute values during normalization and standardization because of the bipolar nature of the hedonic data.

The data were analyzed using repeated-measures analyses of variance (ANOVAs) followed by Tukey’s honestly significant difference tests for post-hoc examination of specific contrasts and interactions. The normality assumption was tested by the Shapiro–Wilk W test, as well as by a skewness test. All statistical analyses were performed using Statistica 8 (StatSoft, Inc.).

Results

Due to the differences in the nature and distributions of the data obtained from the different scaling methods, statistical comparisons were first made between the LHS and ME. The data from the LHS were then further compared with those obtained using the 9-point hedonic scale.

Labeled hedonic scale versus magnitude estimation

Figure 3 compares the data obtained using the LHS with those obtained with ME. A repeated-measures ANOVA revealed a significant effect of stimulus ($F(25, 1350) = 115.11$, $P < 0.0001$), indicating that both methods yielded differences

Table 4 The list of food items names used in Experiment 2

	Food item names
A	Your favorite ice cream with your favorite topping
B	Chocolate chip cookies
C	Chocolate truffles
D	Your favorite pasta
E	Fresh fruit salad
F	Apples
G	Grilled chicken
H	A piece of warm bread with butter
I	Freshly squeezed orange juice
J	Warm chicken noodle soup
K	Cranberry juice
L	French fries
M	Tortilla chips
N	Peanuts
O	Low calorie salad dressing
P	Brussel sprouts
Q	Chilled tomato juice
R	Cold plain oatmeal
S	Cold lima beans
T	Warm buttermilk
U	Soggy corn flakes
V	Stale white bread
W	Spam
X	Beef liver
Y	Uncooked sardines
Z	Pickled pig's feet

in hedonic ratings across items. More importantly, there was no main effect of scale ($F(1, 54) = 1.03, P = 0.316$), nor was there a scale by stimulus interaction ($F(25, 1350) = 1.48, P = 0.07$). Furthermore, post-hoc tests revealed that none of the food item names were rated significantly different between the 2 scales. These results indicate that the LHS yielded data nearly identical to that obtained using ME, which implies that the semantic descriptors of the LHS did not distort the hedonic ratings (i.e., the spacing of the descriptors are valid) and that the scale produces ratio-level data equivalent to ME.

Labeled hedonic scale versus 9-point hedonic scale

Figure 4 shows the distributions of hedonic ratings for 4 examples of food items obtained from the LHS and the 9-point

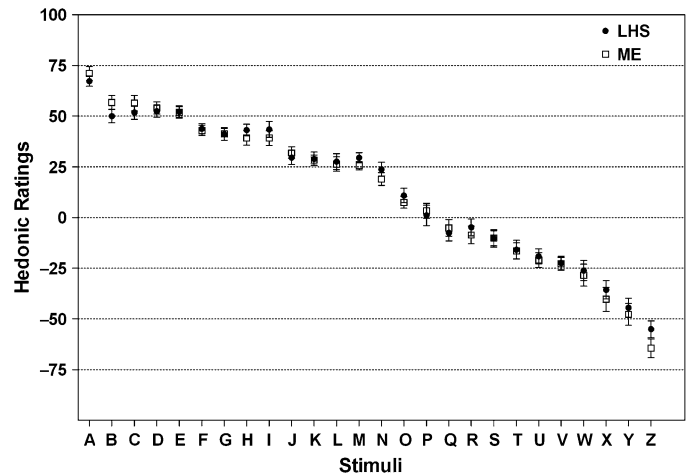


Figure 3 The means \pm standard error of the hedonic ratings for 26 food item names. The means for ME were standardized to the LHS data.

hedonic scale. Because data from the 9-point scale are often not normally distributed (O'Mahony 1982; Gay and Mead 1992; Villanueva et al. 2000, 2005), especially for extremely liked or disliked stimuli (Peryam et al. 1960), it was of interest to determine whether the data from both scales deviated from normality. The Shapiro–Wilk W test showed that the data for 21 of 26 food item names obtained using the LHS were normally distributed. For the other 5 food items (L, M, R, T, and V), further investigation indicated that the distributions were symmetrical, but the peakedness of the distributions was higher than normal (i.e., infrequent extreme deviations), which is considered less problematic because t -tests and ANOVAs are especially robust to those deviations (Miller 1997). In contrast, significant departures from normality were seen in data for all items obtained using the 9-point hedonic scale ($P < 0.01$). The deviation from normality was attributable to 1) a ceiling effect, which resulted in highly skewed data for “extremely liked” or “extremely disliked” items; and 2) the categorical nature of the 9-point scale, which resulted in rating distributions that were either flatter or more peaked than the normal distributions.

ANOVAs conducted on the hedonic ratings of the food item names obtained from the 2 scales showed that both scales yielded a significant effect of stimulus ($P < 0.00001$). Yet, comparisons of the mean ratings of all possible pairs of food items indicated that the LHS afforded slightly better discrimination among stimuli than did the 9-point hedonic scale: Ratings were significantly different for 228 versus 215 of the 325 possible pairs for the LHS and the 9-point hedonic scale, respectively.

As mentioned earlier, the 9-point hedonic scale is a category scale in which psychologically unequal distances are delineated by equally spaced descriptors. This disagreement between scale structure and the hedonic continuum, which the scale is intended to measure, renders the meaning of

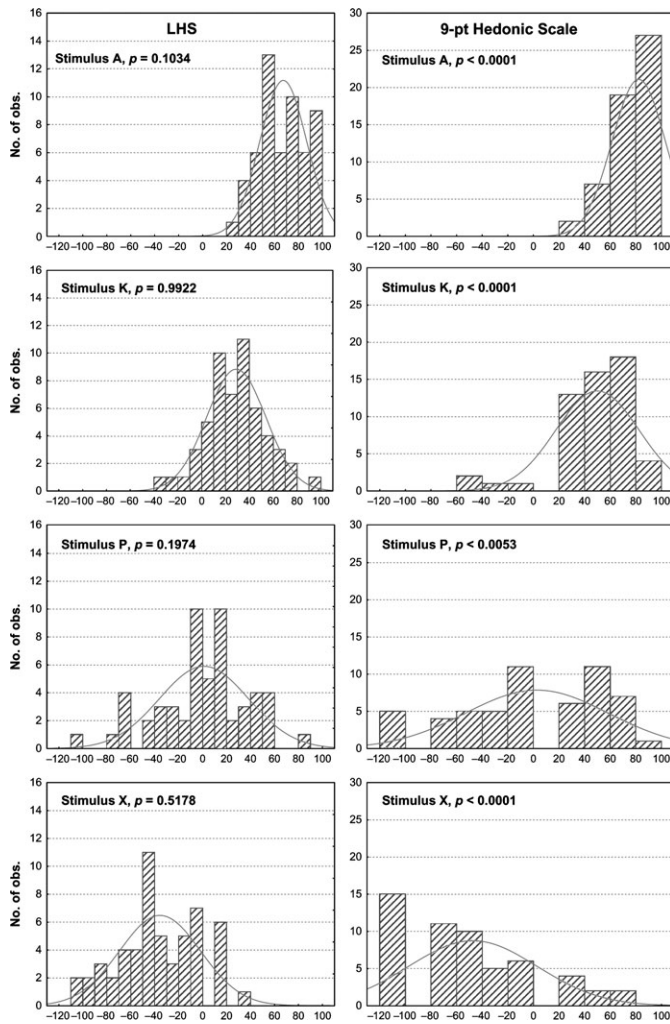


Figure 4 Histograms of hedonic ratings compared with an expected normal distribution for the 2 scales. The data are for items that were rated “like extremely,” “like moderately,” “neutral,” and “dislike very much” items, respectively. The P -values denote the Shapiro–Wilk W test results.

averaged ratings ambiguous. Even so, since its development (Peryam and Girardot 1952), it has been common practice to use mean hedonic scores from the 9-point scale to make general predictions about the acceptance level of foods. Hedonic ratings obtained with the 2 scales were therefore compared directly by placing the data from the 9-point hedonic scale at their equivalent locations relative to the descriptors on the LHS (Figure 5). This was done by calculating the position of each mean rating on the 9-point hedonic scale as a percentage of the distance between adjacent descriptors, then placing the data point at the same percentage of the distance between the 2 categories on the LHS. For example, a mean rating that was midway between “like very much” and “like extremely” on the 9-point scale (i.e., 8.5) was placed midway between the same 2 descriptors on the LHS. The semantic ratings obtained with the 2 scales differed substantially for nearly all food items except those that were neither liked

nor disliked (i.e., had average ratings near neutral), with the 9-point hedonic scale tending to underestimate the degree of liking and disliking of most items compared with the LHS. Although the total possible range of hedonic ratings is less on the 9-point scale because it is bounded by the categories “like extremely” and “dislike extremely,” the range of hedonic ratings was also smaller for the categories that are shared by both scales. This result was not unexpected given the difference in scale structures, though degree of compression in the 9-point scale ratings, which for some items extended across an entire semantic category, was surprisingly large. In summary, compared with the 9-point hedonic scale, the LHS yielded data that were slightly more discriminative, satisfied the normality assumption for parametric statistical analysis, and were much more resistant to end effects.

Discussion

The results of the present study indicate that the LHS shares 2 fundamental properties with the gLMS (Green et al. 1993, 1996; Bartoshuk et al. 2002). Because the locations of its semantic labels were determined by direct estimation of their perceptual and psychological magnitude within the context of a wide range of remembered and imagined experiences, the LHS can be assumed to provide 1) ratio-level data equivalent to that produced by ME (Stevens 1971) and 2) meaningful semantic information. The assumption of ratio-level data was supported by the evidence that the hedonic magnitudes of 26 food item names measured using the LHS and ME were virtually identical (Figure 3). This result, which was also found for comparisons of intensity ratings made with ME and the original LMS (Green et al. 1993), confirms that the presence of the semantic descriptors on the LHS in no way distorts the hedonic ratings and thus that their spacing is valid. Invalid descriptor placement would skew the hedonic ratings by drawing them toward incorrect locations on the scale, such that the results would deviate from those obtained in the unstructured ME task. Valid label placement also allows use of semantic information to describe the degree of liking or disliking of stimulus items in terms of the location of their mean ratings on the LHS. For example, an item could be meaningfully described as being “liked slightly” or “liked a little less than moderately.”

The LHS also has distinctive features that derive from the inherent nature of the hedonic dimension of perception. The most obvious of these is that to enable ratings of like and dislike to be made in a continuous manner, it is constructed as a bipolar scale, with neutral in the center. Importantly, our data show that the positive and negative descriptors are statistically symmetrical around neutral, which agrees in general with other affective scales (Schutz and Cardello 2001; Guest et al. 2007). A less obvious but notable difference is that data obtained with the LHS are distributed differently across subjects than data obtained with the gLMS.

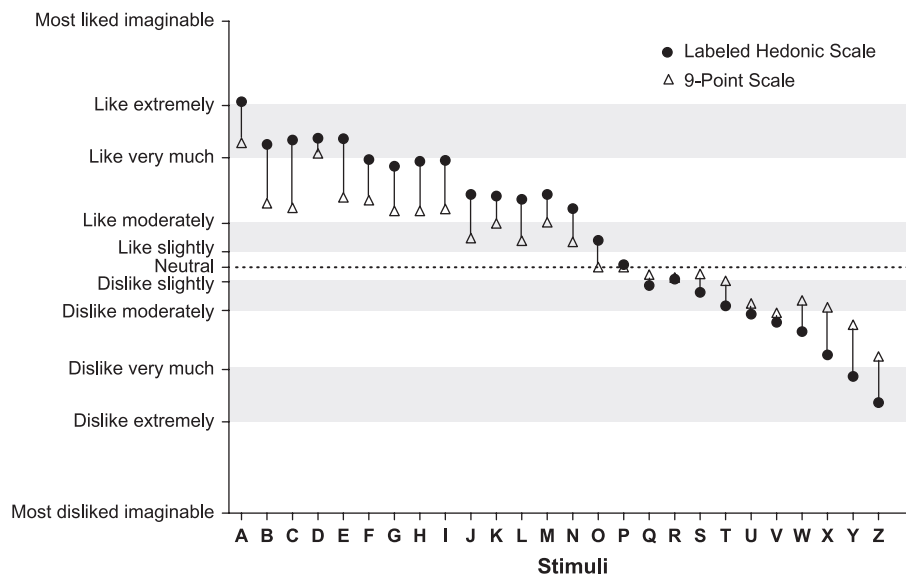


Figure 5 The mean hedonic ratings for the 26 food items from the LHS and the 9-point hedonic scale. The means for the 9-point hedonic scale were adjusted to be plotted on the LHS (see text for explanation).

Although intensity ratings collected via the gLMS (or via ME) approximate a log-normal distribution (Marks 1974; Green et al. 1993), ratings of liking and disliking derived from the LHS are normally distributed. Consequently, the arithmetic mean rather than the geometric mean (which is commonly used for the gLMS and ME) should be used to describe the central tendency of liking and disliking across subjects. The reason for this difference in response distribution is unclear, but it must arise from fundamental differences in the biological and experiential factors that underlie individual differences in the perception of sensation intensity versus those that underlie liking.

Like other category-ratio scales (Borg 1982; Green et al. 1993), an advantage of the LHS is its ability to evaluate hedonic differences across individuals and groups. To some extent, any labeled category scale can be used to make general predictions about the liking of foods or flavors (Peryam and Girardot 1952). It is inadvisable, however, to compare hedonic responses across individuals or groups unless a scale has a frame of reference that can be assumed to be common to all individuals. For the LHS, its end anchors of “most liked sensation imaginable” and “most disliked sensation imaginable” create a semantic framework that encompasses the full range of imaginable hedonic experiences. In theory, this framework is comparable for all individuals who share similar sensory experiences throughout life. However, sources of differences in hedonic reactivity can never be ruled out, nor can they be mitigated by the use of standard stimuli or training, as the effects of these strategies would be influenced by the same idiosyncratic factors. For example, no standard stimulus exists that can be safely assumed to be liked equally by everyone. With this caveat in mind, placing hedonic ratings in the broadest and most natural context has the greatest

potential to provide meaningful data on the average liking or disliking of specific stimuli and on the relative hedonic perceptions of individuals and groups.

Critical to constructing this framework is the use of a rich and varied hedonic context during the semantic scaling task. The frame of reference for the LHS was established by measuring the hedonic magnitudes of the scale’s 10 descriptors (excluding neutral) in the context of wide-ranging experienced and imagined sensations (Table 2). This scale development strategy sets the LHS apart from other labeled affective scales. For example, although the LAM scale (Schutz and Cardello 2001) was developed for the purpose of assessing food preference and was created using subjects experienced in such tasks, no examples of foods or other hedonic stimuli were included within the scale development task itself. In addition, the OPUS was developed in a context that was dominated by examples of the perceived intensities of various oral sensations rather than wide-ranging examples of hedonic stimuli (Guest et al. 2007).

Some concern has been expressed that the use of a wide hedonic context might result in the loss of resolution and discriminating power toward the lower end of a scale (Cardello et al. 2008). This concern, which was raised with respect to discriminating differences in liking and disliking of foods, is based on the assumption that the maximum liking/disliking of flavors or foods may never approach the most liked/disliked experience of any kind. Although this assumption is debatable (most people would agree that their favorite foods and beverages impart exceptionally pleasurable experiences) and deserves to be studied, there is no evidence that a wider hedonic range does in fact impair the overall sensitivity of a scale. On the contrary, data from Cardello et al. (2008) provided indirect evidence that using extreme anchors (i.e.,

greatest imaginable liking/disliking for experiences of any kind vs. for foods and beverages) produced only a small tendency toward compression that was not statistically significant. The same experiment further revealed that acceptance ratings of food products were differentiated equally well with the 2 different end anchors. Finally, Figures 3 and 5 of the present paper show that near neutral, where the loss in resolution would be expected to be greatest, ratings obtained with the LHS do not deviate significantly from those produced by either ME or the 9-point scale. Overall, these results demonstrate that the advantages afforded by a broad frame of reference do not come at the cost of a loss of sensitivity or validity.

Indeed, the results of Experiment 2 showed that the LHS, with its expanded frame of reference, yielded data with slightly greater sensitivity than the traditional 9-point hedonic scale for differentiating affective responses to 26 food items. This finding is not surprising, because the categorical nature (Parducci and Wedell 1986) and fewer response choices of the traditional 9-point hedonic scale often result in a lack of discriminability of individual preferences (Simone and Pangborn 1957; Moskowitz 1982; Marchisano et al. 2003; Villegas-Ruiz et al. 2008). The fact that ratings are made on a continuum rather than categorically also enables the LHS to reveal gradations in liking/disliking for stimuli that would otherwise fall into the same category. This advantage is particularly notable for stimuli or sensations that are more than extremely liked or disliked, where the 2 additional labels at the ends of the LHS (i.e., most liked/disliked sensation imaginable) allow subjects to express differences in hedonic magnitude among extremely liked and extremely disliked items.

Just as important, the all-inclusive end points of LHS make the scale less vulnerable to ceiling or end effects. As pointed out in the Introduction, the 9-point hedonic scale is highly vulnerable to this common form of response bias. Because subjects often tend to be conservative in their use of extreme categories (Hollingworth 1910; Stevens and Galanter 1957; Moskowitz 1982; Parducci and Wedell 1986), the 9-point scale is effectively truncated to a 7-point scale for many subjects, which further reduces the ability to detect differences among extremely liked or disliked sensations. This effect is evident in the current data which show that only 2 food items were rated more than “liked very much” on the 9-point hedonic scale, whereas on the LHS, 1 item was rated more than “liked extremely” and 4 items were rated between “liked extremely” and “liked very much” (Figure 5). An important byproduct of the ceiling effect is that it skews the distribution of responses, which can lead to violations of the normality assumption for parametric statistical analyses and thus to the necessity of using less powerful nonparametric analyses or testing a larger number of subjects to approximate normality. In contrast, ratings obtained with the LHS showed it to be much less vulnerable to this type of response bias (Figure 5), and as a consequence, its data distributions did not deviate from normality (e.g., Figure 4, the top left).

It is important to point out that the LAM scale of Schutz and Cardello (2001) shares with the LHS the advantages of reduced ceiling effects and less-skewed data. However, as can be seen in Figure 2, the LAM scale and the LHS are markedly different scales with quite different semantic structures. Most notable was the finding that “moderately” lies much nearer the midpoints (± 50) of the positive and negative hedonic ranges on the LAM scale than it does on the LHS. In addition, similar differences in locations of the other semantic labels result in very different ratios in hedonic magnitudes between the LHS and the LAM scale. For example, the ratio between hedonic ratings of “like slightly” and “like extremely” on the LAM scale is only 6.6–1, compared with 10.5–1 on the LHS. The smaller range of hedonic magnitudes on the LAM scale, and the placement of “moderate” nearer the middle of its positive and negative hedonic ranges, are both consistent with the use of subjects in the scale development task who were experienced with traditional category scaling (where moderate lies in the middle of the positive and negative hedonic ranges) but not with ME. The broad hedonic context that was generated in the LHS task may also have contributed to the different mean ratings of the semantic labels.

Perceptual context may also be a factor in the differences between the LHS and the OPUS (Guest et al. 2007), which closely resembles the LAM scale. Because subjects who participated in the OPUS study were given practice and experience with ME, the differences from the LHS and similarities to the LAM scale must be due to factors other than subject training. A more likely factor was the semantic context in which the OPUS was developed, which, as alluded to above, was qualitatively narrow (oral sensation only), did not include painful sensations (Green et al. 1996), and contained mainly intensity examples. In addition, subjects in the OPUS experiment made ratings of descriptors of wetness and dryness within the same perceptual context in which they rated hedonic descriptors, albeit in separate blocks of trials. It is perhaps not surprising that under these circumstances, the scales derived for the 2 different perceptual domains had nearly identical semantic structures with “moderate” close to the middle.

It is noteworthy that moderate, the only intensity descriptor of the bipolar gLMS that has a semantic counterpart on the hedonic scales, occupies nearly the same location as “moderately liked/disliked” on the LHS (Figure 2). This close correspondence is consistent with the evidence from Moskowitz and Chandler (1977) that the perceptual domains of intensity and affect have similar (nonlinear) scalar characteristics. To the extent that this is true, the bipolar gLMS has the potential to yield data similar to that obtained with the LHS. However, realization of this potential depends critically on whether the other intensity descriptors of the gLMS, such as “barely detectable” and “strong,” occupy locations on the scale that are valid in terms of hedonic magnitude as well as perceived intensity. To date, no studies have been published that validate the bipolar gLMS as a hedonic category-ratio scale that is capable of providing data comparable to ME.

In summary, the present data show that the LHS provides significant quantitative, semantic, and statistical advantages over what is still the most widely used method of hedonic measurement, the standard 9-point category scale. By providing subjects with instruction and practice in ME prior to the semantic scaling task, and by establishing the broadest possible hedonic framework in which to perform the task, the resulting scale was found to produce ratio-level data equivalent to ME and so to provide valid semantic information about degree of liking and disliking. These findings indicate that the LHS is a unique psychophysical tool that enables hedonic measurement of tastes, flavors, and foods within the broadest possible context of sensations and experiences. Its broad semantic context also enables the scale to be used to study individual and group differences in hedonic perception with the same confidence and caveats (Bartoshuk et al. 2002) that accompany the use of the gLMS to study group and individual difference in intensity perception. Studies are continuing to further assess the psychophysical properties of the LHS (e.g., response distribution, discriminability, compression effects) with taste and flavor stimuli, and experiments are planned that will determine the ways and the extents to which data acquired with the LHS differs from that obtained by other affective scales that have different hedonic ranges and/or somewhat different semantic structures (e.g., the LAM scale; the bipolar gLMS).

Funding

Oregon State University start-up funds and the National Institutes of Health (RO1 DC05009).

References

- Anderson NH. 1982. Cognitive algebra and social psychophysics. In: Wegener B, editor. *Social attitudes and psychophysical measurement*. Hillsdale (NJ): Lawrence Erlbaum associates. p. 123–148.
- Attneave F. 1962. Perception and related areas. In: Koch S, editor. *Psychology: a study of a science*. New York: McGraw-Hill.
- Bartoshuk LM, Duffy VB, Chapo AK, Fast K, Yiee JH, Hoffman HJ, Ko C-W, Snyder DJ. 2004. From psychophysics to the clinic: missteps and advances. *Food Qual Pref.* 15:617–632.
- Bartoshuk LM, Duffy VB, Fast K, Green BG, Prutkin J, Snyder DJ. 2002. Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: what we have learned from genetic variation in taste. *Food Qual Pref.* 14:125–138.
- Bartoshuk LM, Duffy VB, Hayes JE, Moskowitz HR, Snyder DJ. 2006. Psychophysics of sweet and fat perception in obesity: problems, solutions and new perspectives. *Philos Trans R Soc Lond B Biol Sci.* 361:1137–1148.
- Birnbaum MH. 1980. Comparison of two theories of “ratio” and “difference” judgments. *J Exp Psychol.* 109:304–319.
- Bolanowski SJ, Jr. 1987. Contourless stimuli produce binocular brightness summation. *Vision Res.* 27:1943–1951.
- Borg G. 1982. *A category scale with ratio properties for intermodal and interindividual comparisons*. Berlin (Germany): VEB Deutscher Verlag der Wissenschaften.
- Butler G, Poste LM, Wolynetz MS, Agar VE, Larmond E. 1987. Alternative analyses of magnitude estimation data. *J Sen Stud.* 2:243–257.
- Cardello A, Lawless HT, Schutz HG. 2008. Effects of extreme anchors and interior label spacing on labeled affective magnitude scales. *Food Qual Pref.* 19:473–480.
- Duffy VB, Fast K, Cohen Z, Chodos E, Bartoshuk LM. 1999. Genetic taste status associates with fat food acceptance and body mass index in adults. *Chem Senses.* 24:545–546.
- Engen T, McBurney DH. 1964. Magnitude and category scales of the pleasantness of odors. *J Exp Psychol.* 68:435–440.
- Gay C, Mead R. 1992. A statistical appraisal of the problem of sensory measurement. *J Sen Stud.* 7:205–228.
- Green BG, Dalton P, Cowart B, Shaffer GS, Rankin K, Higgins J. 1996. Evaluating the ‘labeled magnitude scale’ for measuring sensations of taste and smell. *Chem Senses.* 21:323–334.
- Green BG, Shaffer GS, Gilmore MM. 1993. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chem Senses.* 18:683–702.
- Guest S, Essick G, Patel A, Prajapati R, McGlone F. 2007. Labeled magnitude scales for oral sensations of sweetness, dryness, pleasantness and unpleasantness. *Food Qual Pref.* 18:342–352.
- Hellman RP, Zwillocki JJ. 1964. Loudness function of a 1,000 cps tone in the presence of a masking noise. *J Acoust Soc Am.* 35:856–865.
- Hollingworth HL. 1910. The central tendency of judgment. *J Philos Psych Sci Meth.* 7:461–469.
- Jones LV, Peryam DR, Thurstone LL. 1955. Development of a scale for measuring soldiers’ food preferences. *Food Res.* 20:512–520.
- Jones LV, Thurstone LL. 1955. The psychophysics of semantics: an experimental investigation. *J Appl Psychol.* 39:31–36.
- Lawless HT, Heymann H. 1998. *Sensory evaluation of food: principles and practices*. New York: Chapman & Hall.
- Marchisano C, Lim J, Cho HS, Suh DS, Jeon SY, Kim KO, O’Mahony M. 2003. Consumers report preference when they should not: a cross-cultural study. *J Sen Stud.* 18:487–516.
- Marks LE. 1974. *Sensory processes*. New York (NY): Academic Press.
- Marks LE. 1979. Summation of vibrotactile intensity: an analog to auditory critical bands? *Sens Processes.* 3:188–203.
- Marks LE, Bartoshuk LM. 1979. Ratio scaling of taste intensity by a matching procedure. *Percept Psychophys.* 26:335–339.
- Miller RG. 1997. *Beyond ANOVA: basics of applied statistics*. London: Chapman & Hall.
- Moskowitz HR. 1971. The sweetness and pleasantness of sugars. *Am J Psychol.* 84:387–405.
- Moskowitz HR. 1977. Magnitude estimation: notes on what, how, when, and why to use it. *J Food Qual.* 3:195–227.
- Moskowitz HR. 1982. *Utilitarian benefits of magnitude estimation scaling for testing product acceptability*. Philadelphia (PA): American society for testing and materials.
- Moskowitz HR, Chandler JW. 1977. New uses of magnitude estimation. In: Birch GG, Brennan JG, Parker KJ, editors. *Sensory properties of foods*. London: Applied Science Publishers. p. 189–211.
- O’Mahony M. 1982. Some assumptions and difficulties with common statistics for sensory analysis. *Food Tech.* 36:75–82.
- Parducci A, Wedell D. 1986. The category effect with rating scales: number of categories, number of stimuli, and method of presentation. *J Exp Psychol Hum Percep Perform.* 12:496–516.

- Peryam DR, Girardot NF. 1952. Advanced taste-test method. *Food Eng.* 24:58–61.
- Peryam DR, Pilgrim FJ. 1957. Hedonic scale method of measuring food preference. *Food Tech.* 11:9–14.
- Peryam DR, Polemis BW, Kamen JM, Eindhoven J, Pilgrim FJ. 1960. Food preferences of men in the U.S. armed forces. Chicago (IL): Quartermaster Food and Container Institute for the Armed Forces. p. 1–160.
- Schutz HG, Cardello AV. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *J Sen Stud.* 16:117–159.
- Simone M, Pangborn RM. 1957. Consumer acceptance methodology: one vs two samples. *Food Technol.* 11:25–29.
- Stevens JC. 1957. A comparison of ratio scales for the loudness of white noise and the brightness of white light. [Doctoral dissertation]. [Cambridge (MA)]: Harvard University.
- Stevens SS. 1953. On the birghtness of lights and loudness of sounds. *Science.* 118:576.
- Stevens SS. 1955. The measurement of loudness. *J Acoust Soc Am.* 27:815–829.
- Stevens SS. 1958. Adaptation-level vs the relativity of judgment. *Am J Psychol.* 71:633–646.
- Stevens SS. 1971. Issues in psychophysical measurement. *Psychol Rev.* 78:426–450.
- Stevens SS, Galanter EH. 1957. Ratio scales and category scales for a dozen perceptual continua. *J Exp Psychol.* 54:377–411.
- Villegas-Ruiz X, Angulo O, O'Mahony M. 2008. Hidden and false "preferences" on the structured 9-point hedonic scale. *J Sen Stud.* 23:780–790.
- Villanueva NDM, Petenate AJ, Da Silva MAA. 2000. Performance of three affective methods and diagnosis of the ANOVA model. *Food Qual Pref.* 11:363–370.
- Villanueva NDM, Petenate AJ, Da Silva MAA. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. *Food Qual Pref.* 16:691–703.
- Zwislocki JJ. 1983. Group and individual relations between sensation magnitude and their numerical estimates. *Percept Psychophys.* 33:460–468.

Accepted July 21, 2009