

Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: what we have learned from genetic variation in taste

L.M. Bartoshuk^{a,*}, V.B. Duffy^{a,b}, K. Fast^a, B.G. Green^{a,c}, J. Prutkin^a, D.J. Snyder^a

^a*Yale University School of Medicine, PO Box 208041, New Haven, CT 06520-8041, USA*

^b*School of Allied Health, University of Connecticut, Storrs, CT 06269-2101, USA*

^c*John B. Pierce Foundation Laboratory, 290 Congress Avenue, New Haven, CT 06519, USA*

Abstract

Direct comparisons of sensory or hedonic perceived intensities across individuals are impossible since we cannot share experiences. However, experiences can be compared indirectly if we can identify a standard assumed to be equal, on average, to groups compared. Sensory standards (i.e., magnitude matching) have proved very useful. Intensity adjectives/adverbs have also been used as standards (e.g., This tastes very strong to me; is it very strong to you?). We argue these labels often refer to experiences of different absolute intensity to the groups of interest making the comparisons invalid. An early solution rested on the assumption that the maximum intensity perceivable was equal for all sensory domains and individuals. Comparisons were then possible because all judgments could be made on a zero to maximum scale. We present data showing this assumption is not true.

© 2002 Published by Elsevier Science Ltd.

Keywords: Category scale; Likert scale; Visual analogue scale; Magnitude estimation; Taste; PROP; Across-group comparisons

1. Introduction

Variation in sensory or hedonic experience is relatively easy to measure within individuals. But it is extremely hard to measure across individuals or groups because we cannot share one another's experiences. Two psychophysical traditions have led to attempts to provide comparisons of perceived sensory or hedonic intensities across groups. The older tradition depends on scales of various types labeled with adjective/adverb intensity descriptors (e.g., weak, medium, very strong). The newer tradition derives from the work of S.S. Stevens on direct scaling methods. Both depend on the same logic. In order to make comparisons across groups, we must identify some kind of standard that can be assumed to be, on average, equally intense to each group. When comparisons are made through intensity descriptors, the standards are the descriptors themselves. Unfortunately, as we will discuss below, these often fail to reflect equivalent perceived intensities; this makes these scales invalid for across-group comparisons

(see also Bartoshuk, 2002). On the other hand, the logic underlying comparisons using direct scaling methods has permitted what we believe to be valid across-group comparisons in some circumstances. Although we know of no stimulus likely to produce the same perceived intensity to all individuals, we can identify stimuli likely to have the same perceived intensity, on average, to particular groups (we need only assume that the modalities to be investigated and the standard modality are not related to one another for those groups).

Given the wide reliance on labeled scales, there would be considerable value in attempting to find ways to alter these scales to make them produce valid across-group comparisons. We believe that increasing insight into our use of intensity descriptors may suggest such ways. These can then be tested against scaling that uses sensory standards. A combination of these approaches may provide the most powerful method of all.

2. Adjective/adverb-labeled scales used to compare perceived intensities across subjects/groups

In everyday language we compare sensations using intensity adjectives (e.g., This tastes strong to me. Does it taste strong to you?) along with adverbs to modify the

* Corresponding author. Tel.: +1-203-785-2587; fax: +1-203-737-2245.

E-mail address: linda.bartoshuk@yale.edu (L.M. Bartoshuk).

adjectives (e.g., This tastes very strong to me. Does it taste very strong to you?). The validity of such comparisons depends on the assumption that these intensity descriptors refer to the same absolute intensities to all.

One of the most ancient scales dates to the astronomer Hipparchus (190–120 BC) who classified stars into six categories by their brightness. In the modern era, Likert (1932) used a category scale for the study of attitudes. A category scale in common use in food preference studies is the Natick nine-point scale developed for use by the military in 1949 (e.g., see Jones, Peryam, & Thurstone, 1955; Peryam & Girardot, 1952).

The “graphic rating scale” arose by removing the category divisions (Hayes & Patterson, 1921; Patterson, 1923). Freyd (1923) described the graphic rating scale as a combination of rating on a straight line (a preview of what was to become the visual analogue scale) “under which were printed descriptive phrases indicative of varying degrees of the trait, from one extreme to the other.” Guilford (1954) credits Champney (1941) for the “innovation” of using a vertical rather than a horizontal line.

The issue of whether or not the descriptors on category or graphic rating scales were equally spaced led Lasagna (1960) to ask patients to place pain descriptors (slight, moderate, severe and very severe) at the appropriate places on a vertical pain “thermometer.” He found that they were not equally spaced (a similar conclusion resulted from other pain research using a visual analogue scale (Berry & Huskisson, 1972). This was a preview of the efforts of psychophysicists to place intensity adjectives on ratio scales (see below).

The visual analogue scale (VAS) came into its own in the 1960s (e.g., Aitken, 1969; Aitken, Ferres, & Gedy, 1963; Clarke & Spear, 1964; Silverstone & Stunkard, 1968). For example, to measure the apprehension experienced by fighter pilots in ten defined situations, Aitken asked pilots to rate their apprehension on a 100 mm line labeled “maximal relaxation” at one end and “maximal panic” at the other end. The VAS is now extremely popular in a variety of fields; a search under “Visual Analogue Scale” via PubMed produced 16,812 references from 1967 through 2001. Typically it is described as a line labeled at its ends with the “minimum and the maximum rating” for a particular variable (e.g., see Hetherington & Rolls, 1987).

The relation between category and ratio scales was of considerable interest to S.S. Stevens and his students and colleagues. Ratings of sensory intensity on category scales are generally concave downward when plotted against ratings from a ratio scale (e.g., see Stevens & Galanter, 1957). The removal of the category ratings to make the VAS appears to have given the VAS ratio properties (e.g., see the pain VAS in Price, McGrath, Rafii, & Buckingham, 1983).

3. Labeled scales and erroneous across-group comparisons

Adjective/adverb labeled scales can obviously be used for within-subject comparisons. They can also be used for across-group comparisons when the members of the groups are randomly assigned; this insures that there will be no difference, on average, in the way the groups use the labels. However, these scales have come to be used to compare groups that differ in systematic ways (e.g., sex, age, clinical status). This usage rests on the implicit assumption that the intensity labels refer to the same absolute perceived intensity, on average, for each group. Reflection shows this assumption is false. S.S. Stevens (1958) made this point more than 40 years ago: “Mice may be called large or small, and so may elephants, and it is quite understandable when someone says it was a large mouse that ran up the trunk of the small elephant.” This example illustrates that the absolute intensity indicated by an adjective varies depending on the domain to which the adjective is applied. Intensity adverbs modify adjectives; thus we can describe a very large mouse and a very small elephant (see Cliff, 1959; Mosteller & Youtz, 1990) for accounts of how intensity adverbs modify adjectives).

Even within the same domain, the absolute intensities associated with a given descriptor can vary across individuals depending on differences in experience or physiology. Once we understand this, it is clear that the use of labeled scales to make across-group comparisons will be invalid whenever the members of the groups differ systematically in experience or physiology. It is important to note that the reason most investigators initiate a study on group differences is that they suspect the differences exist. But if the differences do exist, then this provides the very circumstances that could make the absolute intensities to which the scale labels refer vary across the groups.

Some of the early developers of various labeled scales were aware of their limitations. In an early paper on the VAS, for example, Aitken (1969) noted, “The same word used by different people need not convey that they experience the same feeling, neither does comparable positioning of marks on lines.” This was not always appreciated in the years to come. One can imagine a variety of group differences that could involve systematically different experiences in some domains: sex, race, age, clinical status (e.g., alcoholism, obesity, psychiatric diagnoses, etc.). The problem has been particularly troublesome in studies on genetic variation in taste.

It is important to note that psychophysicists have long known ratings of perceived intensity (however they are obtained) can be altered by context and experience (e.g., see Gescheider, 1988; Helson, 1964; Parducci, 1965; Zoeke & Sarris, 1983). Although these effects were generally demonstrated by within-subject comparisons (e.g., demonstrating how alterations in context or

experience could change an individual's ratings), the insights gained also apply to across-group comparisons and imply that they would be invalid whenever context or experience varies systematically across the groups.

Experts in a variety of fields have warned specifically of the consequences of invalid across-subject/group comparisons. For example, Narens and Luce (1983) disputed some economists' belief in the "intercomparability of utility" the idea that the magnitude of value can be compared across individuals. They provided an all too familiar illustration of their point by noting that our confidence in assessing others' preferences may lead us to introduce two people that we predict will be compatible only to find ourselves very mistaken. Biernat and Manis (1994) noted that a label could indicate different absolute intensities depending on the target to which it is applied. For example, "very tall" indicates a different height when applied to a woman than when applied to a man. Birnbaum (1999) addressed the absurd results that can occur when comparisons are made across groups using different contexts to make judgments. Birnbaum asked different groups to judge the size of the numbers 9 and 221 on a 10-point scale ranging from "very very small" to "very very large." Because of the different contexts brought with the two numbers, the average scale values for 9 and 221 were 5.13 and 3.10, respectively, leading to the apparent conclusion that $9 > 221$.

Frequency and probability descriptors are also perceived to reflect different magnitudes across subjects/groups (e.g., Hakel, 1968; Kong, Barnett, Mosteller, & Youtz, 1986; Mapes, 1979; Simpson, 1944; Wallsten & Budescu, 1995; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). Woody Allen and Marshall Brickman immortalized this in their 1977 movie "Annie Hall." When Alvy (Allen) and Annie (Diane Keaton) in split screen visits with their respective psychiatrists describe the frequency with which they have sex, Alvy says "Hardly ever," and Annie says "Constantly" and then both add, "three times a week."

We will show how invalid comparisons across groups will most often blunt the comparisons; that is, the differences across groups will seem smaller than they are in reality. However, in some cases, differences that seem to be real will actually be in the wrong direction. We have identified examples of this in studies on genetic variation in taste (Bartoshuk, Duffy, Fast, Green, & Snyder, 2002).

4. Stimuli as standards: magnitude matching

S.S. Stevens (1956, 1957) revolutionized psychophysics in the 1950s by introducing direct scaling methods. The most popular of these, magnitude estimation, directed subjects to assign numbers to sensations such that one sensation twice as intense as another is

assigned a number twice as large, etc. Stevens devised this method to produce scales with ratio properties, a highly desirable attribute. However, it is important to note that magnitude estimate data can only show the rate at which perceived sensory intensities grow with increases in the stimulus; absolute magnitude estimates cannot be compared across subjects/groups (see Michell, 1999, for a critique of measurement as Stevens saw it).

5. Genetic variation in taste

Fox (1931) discovered taste blindness to PTC (phenylthiocarbamide). About 25% of the US population cannot taste PTC (nontasters); the other 75% taste it as bitter. Family studies (Blakeslee, 1932; Snyder, 1931) suggested that nontasting was a simple Mendelian recessive. Very early in the history of PTC studies, threshold measures became the gold standard (Harris & Kalmus, 1949) [See Fernberger (1932) for a rare supra-threshold study from that era]. The PTC threshold distribution was essentially bimodal; most subjects had thresholds that fell clearly into either the nontaster distribution (higher concentrations of PTC) or the taster distribution (lower concentrations of PTC). However, Kalmus (1958) showed that the average PTC threshold for tasters with only nontaster siblings was slightly higher than that for tasters with only taster siblings; this suggested that PTC tasting shows incomplete dominance (also shown by Reed and her co-workers (Reed, Bartoshuk, Duffy, Marino, & Price 1995). However, the taster PTC threshold distribution did not clearly separate into two groups; thresholds could not sort the heterozygous from the homozygous dominant tasters.

In 1975, my lab wanted to scale the perceived bitterness of PTC with magnitude estimation. However, we faced the problem noted above: magnitude estimates cannot be compared across subjects/groups. The early PTC studies tested a variety of compounds and concluded that only those containing the $N-C=S$ group produced the bimodal distribution associated with PTC. Thus, we reasoned, if we asked subjects to rate PTC relative to another compound (acting as a standard) that did not contain the $N-C=S$ group, we could make absolute comparisons of the average bitterness perceived by nontasters and tasters. Using NaCl as our standard, we found that caffeine, a bitter compound not containing the $N-C=S$ group, was less bitter to nontasters than to tasters (Hall, Bartoshuk, Cain, & Stevens, 1975).

Subsequently, we switched from PTC to PROP (6-n-propylthiouracil). PROP is a medication and thus safety information is available to protect human subjects; in addition, PROP lacks the sulfurous odor associated with PTC (Fischer, 1967; Lawless, 1980). Studies using

PROP led to new findings. First, some sweet compounds (that clearly did not contain the N–C=S group) tasted more intense to tasters than to nontasters; these included saccharin (Bartoshuk, 1979) and sucrose (Gent & Bartoshuk, 1983) as well as sodium and potassium benzoate (Bartoshuk, Rifkin, Marks, & Hooper, 1988). Some milk products tasted more bitter to tasters (Marino et al., 1991). In addition, the greater intensities perceived by tasters extended to a non-taste oral sensation: the burn of capsaicin (chilis) (Karrer & Bartoshuk, 1991).

As data accumulated, it became clear that some tasters perceived much greater bitterness from PROP than did others. These supertasters (Bartoshuk, 1991) tended to experience the most intense sensations not only from a variety of other taste stimuli but also from oral irritants and oral thickeners including fats (Bartoshuk, 1993a, 1993b; Bartoshuk et al., 1992, 1993; Duffy, Bartoshuk, Lucchina, Snyder, & Tym, 1996; Phillips, Bartoshuk, Peterson, & Duffy, 2001; Prescott, Johnstone, & Munro, 2001; Prescott & Swain-Campbell, 2000; Prutkin, Fast, Lucchina, & Bartoshuk, 1998; Snyder, Lucchina, Duffy, & Bartoshuk, 1996; Tepper & Nurse, 1997).

The fact that tasting was not related solely to the N–C=S group in a specific set of bitter compounds made us concerned about the use of NaCl as a standard. Fortunately, J.C. Stevens had pioneered the study of cross-modality matching (Stevens, 1959; Stevens & Marks, 1965). His work suggested that subjects can match perceived intensities even though stimuli may vary in quality; J.C. Stevens and Marks formalized the idea of a standard from a modality unrelated to the modality under study as the “method of magnitude matching” (Marks & Stevens, 1980; Stevens & Marks, 1980). Thus we could use non-taste standards in taste studies (Marks et al., 1988). Using a sound standard (assuming that nontasters, medium tasters and supertasters hear the same, on average) revealed that the saltiness of NaCl varies with taster status: nontasters perceive the least and supertasters the most saltiness (Bartoshuk, Duffy, Lucchina, Prutkin, & Fast, 1998). This suggests that the magnitude of PROP effects will be larger when tested with a sound standard [e.g., see Ko et al. (2000) for taste data and Prutkin et al. (1998), for data on other oral sensations].

Interestingly, the cross-modality matching work of J.C. Stevens had an early impact on the pain field. Peck, a physician in private practice, wanted to measure pain (Peck, 1967) and anxiety (Peck, 1966). He asked subjects to turn the knob of an audiometer until the sound intensity matched that of the pain or anxiety experienced, presaging the formal development of magnitude matching by many years. Another matching technique preceded even Peck’s work. In 1962, Kast used mechanically mediated pressure to the leg to create pain

that could be varied in intensity by a patient suffering pathologic pain. The patient was instructed to press a button to increase the pain to the leg and to release the button when the pain matched the pathologic pain.

6. Magnitude estimation and across-group comparisons: Borg/Teghtsoonian range theory

Stevens was interested in comparing sensory modalities; he was not interested in making comparisons across subjects or groups. Thus he was not concerned that magnitude estimates cannot be compared across subjects. Borg (1961), who was concerned and sought solutions to this problem, suggested that the range from no sensation to maximal sensation within sensory modalities was equal for all subjects. He initially cautioned that this assumption might not be “suitable for all sense modalities.” However, Teghtsoonian (1973) argued that “for all perceptual continua the maximum range of subjective magnitude is constant.” In 1992, Borg described a pilot study he had conducted in 1985 at a conference of psychophysicists (Borg, 1992). He asked the attendees to “estimate conceived maxima for different modalities.” He reported, “All of the psychophysicists gave the highest estimates to pain, then came a big class consisting of loudness, brightness, perceived effort and exertion etc., in which the maxima were about the same. The lowest estimates—the ‘weakest modality’—was taste, especially the quality of sweetness.” Thus he argued (Borg, 1994), “...maximal intensities are about the same for most sensory modalities (except for pain)....” By this view, across subject/group comparisons would become logically possible. Subjects would only have to rate their sensations proportionately on a scale from zero to maximum. All such ratings across subjects and across sensory continua would then be comparable with the exception of pain. (We will challenge that view in the present study.)

Borg also felt that subjects would be aided by having adjective/adverb descriptors like “weak,” “very strong,” etc. on the scale. He tested various spacings to try to give his scale ratio properties but he did not use empirical methods to place the adjectives (e.g., see Borg, 1982).

Moskowitz (see Moskowitz, 1977; Moskowitz & Chandler, 1977) located intensity labels (barely, slightly, moderately, very, extremely) empirically for foods/beverages. He did this in the following way. At the end of a standard magnitude estimation experiment, he asked subjects to provide magnitude estimates of the intensity labels as if those labels referred to stimuli of the same type as those just experienced. Of special importance, he did this for a variety of attributes (sweetness, flavor intensity, consistency, color deepness and liking/disliking) in studies with various food products. The relative

spacing of the labels remained essentially the same (Moskowitz & Chandler, 1977) and was similar to that proposed by Borg. This was the first time that the spacing of intensity adjectives was determined empirically and compared for a variety of sensory domains as well as for a hedonic domain.

Other studies related intensity labels to pain experiences. Gracely, McGrath, and Dubner (1978) asked subjects to provide magnitude estimates of pain intensity descriptors (e.g., weak, moderate, strong) as well as affective terms (e.g., unpleasant, awful, unbearable). Heft and Parker (1984) associated pain intensity descriptors (faint, weak, mild, moderate, strong, intense) with electrocutaneous shock to the wrist. Karer and Bartoshuk (1991) followed the Moskowitz strategy and associated intensity descriptors (very weak, weak, medium, strong, very strong) with the oral burn produced by capsaicin.

Up to this time, only Borg had anchored the top of the scale to “maximal.” Green, Shaffer, & Gilmore (1993) asked subjects to provide magnitude estimates for a variety of commonly experienced oral stimuli (including oral pain) along with estimates for intensity descriptors (as the earlier investigators had done) but, in addition, they asked subjects to locate the top of the scale: “strongest imaginable sensation.” Thus the Labeled Magnitude Scale (LMS) devised by Green and colleagues was the first to be both empirically derived and anchored at the top with the most intense sensations possible for the domain under study. Green and colleagues specifically limited their scale to oral sensations because of concern that the strongest possible sensation might vary across sensory modalities.

Schutz and Cardello (2001) constructed a scale in a fashion similar to that used by Green et al; the Schutz–Cardello scale was derived for hedonic experiences associated with foods. Although the descriptors in the studies vary, to the extent that they can be compared, the spacings are remarkably similar across studies; the scales have a logarithmic appearance because the spaces between descriptors tend to get larger as we move from weak to stronger descriptors. We should expect a certain amount of variability in the spacings since a variety of “biases” are known to affect psychophysical judgments (e.g., the relative number of descriptors reflecting the weakest versus the strongest experiences, etc.; see Gescheider, 1988). However, the overall agreement across the scales is impressive, nonetheless. We suggest that the similarity of these scales is the result of the way we learn to use intensity adjectives/adverbs. The spacing is relatively invariant because it reflects the relations among the descriptors and these remain essentially the same independent of the domain to which they are applied. We can imagine a family of labeled ratio scales stretched or compressed to fit the domains of interest. Although the existing data support this conclusion, it

would be of interest to test this directly using the same subjects to locate intensity descriptors in various domains.

7. General Labeled Magnitude Scale (gLMS)

It is important to note that the addition of adjective/adverb descriptors to ratio scales does not solve the problem of the relativity of the descriptors across context and experience. When Green and his colleagues constructed the LMS, they anchored the top of the scale to oral sensations including pain assuming that oral pain would be equivalent, on average, across groups of subjects. However, the intensities of sensations produced by irritants placed on the tongue are linked to tongue anatomy. Taste buds are surrounded by nerve fibers believed to mediate burn/pain (Finger, Nelson, Bryant, & Moore, 1994; Nagy, Goedert, Hunt, & Bond, 1982; Silver & Finger, 1991; Whitehead, Beeman, & Kinsella, 1985) and the number of taste buds is related to PROP status (Bartoshuk, Duffy, & Miller, 1994; Miller & Reedy, 1990; Tepper & Nurse, 1997). If the perception of burn determined maximal oral sensation, the top of the LMS would not be equivalent for non-tasters, medium tasters and supertasters.

The first indication that the top of the LMS might differ across PROP groups came from a study by Snyder et al. (1996) on capsaicin burn and PROP status. Snyder asked subjects to rate the saltiness of NaCl, the bitterness of PROP and the oral burn of capsaicin. Including NaCl permitted the data to be analyzed as a magnitude matching study; ratings were normalized to NaCl. Note that we now know that this would underestimate the magnitude of the PROP effect; yet, normalizing the data to NaCl produced a larger PROP effect than did the raw LMS data. Thus, following Snyder's study, we anchored the top of the LMS with “strongest imaginable sensation of any kind” (e.g., Bartoshuk, 2000; Prutkin et al., 2000). We now refer to this generalized version of the LMS as the gLMS (Bartoshuk, Duffy, Fast, Green, & Snyder, 2001; Bartoshuk et al., in press; Fast, Green, Snyder, & Bartoshuk, 2001; Snyder, Duffy, Fast, Hoffman, et al., 2001; Snyder, Duffy, Fast, Weiffenbach, & Bartoshuk, 2001). Note that Lawless and his colleagues (Lawless, Horne, & Spiers, 2000) have expressed similar concern with use of the LMS to make “inter-group comparisons . . . between groups with different bitter sensitivities.”

We can think of the gLMS as a labeled scale with ratio properties “stretched” to the maximum possible. This would not be identical to the scale proposed by Borg because Borg took the position that the maxima across most modalities are the same and we do not (it is one purpose of the present study to demonstrate that maxima across modalities show considerable variation).

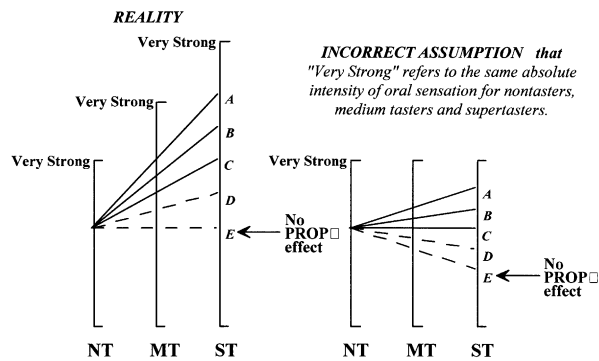


Fig. 1. The left side illustrates the actual differences across the taste worlds of nontasters (NTs), medium tasters (MTs) and supertasters (STs). The right side illustrates the consequences of assuming incorrectly that all three groups live in the same taste world (Bartoshuk et al., in press).

Comparisons of the data collected by Snyder (original LMS) with data collected later by Prutkin (gLMS) revealed the severity of the consequences of making an erroneous assumption about the equivalence of the top of the LMS to all. Plotting the perceived saltiness of NaCl versus the perceived bitterness of PROP produced a positive correlation with the gLMS but a negative correlation with the original LMS (Bartoshuk et al., 2000). We illustrate the way in which a significant positive correlation can appear to be a significant negative correlation in Fig. 1.

Although the example in Fig. 1 is in terms of oral sensations, the principles are the same in any area in which subjects have systematically different experiences so that intensity descriptors will indicate different absolute intensities to different groups. The left side of Fig. 1 suggests the differences in oral sensations across nontasters, medium tasters and supertasters. These were estimated from laboratory studies using magnitude matching (sound standard) with adjectives rated at the end (Bartoshuk et al., 2000). The functions labeled A through E are drawn to reflect differences in the size of PROP effect. For example, function A could reflect a very large effect like that produced by quinine (bitterness), functions B through D could reflect smaller effects like those for sucrose or NaCl (Ko et al., 2000). Function E reflects a sensation that is not correlated with PROP (e.g., oral burn from a non-tongue location; see Karrer et al., 1992). The right side of Fig. 1 shows what happens under the erroneous assumption that “very strong” indicates the same absolute intensity to all. For the largest effects (A and B), assuming that “very strong” is the same to all reduces the size of the PROP effect. However, for C, the PROP effect is completely lost. Note that D and E now erroneously appear to suggest that nontasters actually perceive more intense sensations than do medium and supertasters: a reversal artifact. D in Fig. 1 explains the apparent negative

correlation between PROP bitterness and NaCl saltiness in the Snyder study noted above. We have identified some instances in which the reversal artifact appears to have occurred in PROP studies (Bartoshuk et al., in press).

Comparing the gLMS with magnitude matching (Bartoshuk et al., 2000) revealed that both produce about the same separation across nontasters, medium tasters and supertasters; this strengthens the idea that the assumptions of both procedures are roughly correct. That is, the gLMS may produce valid comparisons across PROP groups and hearing is not closely related to tasting (but see the discussion).

The discovery that perceived taste intensities are associated with density of fungiform papillae provides a tool for evaluating the power of scales to provide valid comparisons across nontasters, medium tasters and supertasters. For example, in a pilot study the gLMS and magnitude matching produced a significant correlation between fungiform papillae density and the bitterness of quinine; two visual analogue scales did not (Prutkin et al., 2000).

Note that we need not use the gLMS to make valid comparisons across nontasters, medium tasters and supertasters. We need only anchor the top of the scale to a domain not related to taste. However, if we selected a domain that was smaller than the taste domain to supertasters, we would produce a ceiling effect: the supertasters would be unable to appropriately locate their taste sensations on such a scale. The gLMS avoids this. To the extent that the “strongest imaginable sensation of any kind” is similar across individuals, the gLMS would be a universal sensory ruler. It is clearly unlikely that this is true. However, whenever “strongest imaginable sensation of any kind” is not systematically different across groups, the gLMS will produce valid comparisons across those groups.

Since Moskowitz and Chandler (1977) found that the spacing of hedonic intensity labels was similar to that for sensory intensity labels (suggesting that hedonic experiences have the same quantification properties as sensory experiences), we created a bipolar hedonic scale using the adjective spacing found by Green et al. (1993). With neutral in the center, negative hedonic ratings were placed on the left and positive hedonic ratings, on the right. The ends were anchored with “strongest imaginable hedonic sensation of any kind” (Bartoshuk et al., 1999; Bartoshuk et al., in press; Chapo, Bartoshuk, Peterson, Phillips, & Duffy, 2001; Chapo, Phillips, Ilich, & Duffy, 2001; Duffy, Fast, Cohen, Chodos, & Bartoshuk, 1999; Duffy, Phillips, Peterson, & Bartoshuk, 2001; Duffy & Peterson, 2000; Peterson & Duffy, 2000a, 2000b; Fast, Duffy, & Bartoshuk, in press; Peterson, Bartoshuk, & Duffy, 1999; Phillips et al., 2001; Snyder, Duffy, Fast Hoffman et al., 2001; Reed et al., 2002; Snyder, Duffy, Fast, Weiffenbach et al., 2001).

The hedonic food scale that Schutz and Cardello (2001) constructed used the same methods as those used by Green et al. (1993). The fact that they found similar spacing among the descriptors further supports the idea that hedonic and sensory intensities show similar intensity properties. We suggest that anchoring the ends of hedonic scales in terms of all hedonic experience provides the same advantage for hedonic scales as it does for sensory scales: it permits comparisons across groups that may live in very different hedonic food worlds.

The data from the experiment presented here will show that the maximum sensations in different modalities are not equally intense; that is, the assumption in range theory that all sensory continua have equivalent maxima is not supported. In addition, we show that even using remembered sensations, nontasters, medium tasters and supertasters can be shown to live in different taste worlds. The present study also shows that the most intense remembered oral burn tends to exceed the most intense remembered oral pain for the subjects most responsive to PROP. Thus the most intense oral sensation is not equivalent across nontasters, medium tasters and supertasters; this limits the validity of the original LMS for the measurement of PROP effects.

8. Experiment

Data were collected from attendees (20–79 years old, inclusive) at lectures given by LMB (data collected prior to the lecture). Attendees filled out a questionnaire, tasted a piece of candy rating it for sweetness and tasted PROP paper rating it for bitterness. All ratings were made on the gLMS.

8.1. Questionnaire

We have been using questionnaires to collect data on hedonic responses to foods as well as bitterness of PROP paper (see below) at lectures since 1993. An early version of the questionnaire utilized the Natick nine-point scale both for hedonic ratings of foods (1 = strongly dislike, 5 = neutral, 9 = strongly like) and for the perceived bitterness of PROP paper (0 = no taste, 1 = very weak, 5 = medium, 9 = very strong) (see (Bartoshuk et al., 1994; Bartoshuk, Duffy, Reed, & Williams, 1996; Duffy & Bartoshuk, 1996) for portions of those data). With the realization that the Natick nine-point scale would cause comparisons across nontasters, medium tasters and supertasters to appear erroneously small (Bartoshuk, 2000; Bartoshuk et al., in press; Fast et al., in press; Lucchina et al., 1998; Prutkin et al., 2000), we switched to the gLMS for measures of PROP paper bitterness and food preferences (Bartoshuk et al., 1999; Bartoshuk et al., in press; Duffy et al., 1999; Fast et al., in press; Snyder, Duffy, Fast, Hoffman et al.,

2001; Snyder, Duffy, Fast, Weiffenbach et al., 2001). With some versions of the questionnaire we also evaluated the oral burn of capsaicin (in a candy) (Bartoshuk, Caseria et al., 1996) and the sweetness of Stop & Shop butterscotch disks (Bartoshuk, 2000; Weiffenbach, Duffy, Fast, Cohen, & Bartoshuk 2000). Recently, we added an item asking subjects if they have ever tasted PROP before. Only subjects for whom this was the first experience tasting PROP were included in this analysis.

The questionnaire consists of a single sheet with a list of foods for hedonic evaluation on one side. These hedonic ratings will not be discussed here (see the references above). The other side of the questionnaire contains a list of sensations to be rated for perceived intensity on the gLMS (see Table 1). Some of the remembered sensations were added to the questionnaire earlier than others and some attendees did not rate each item so the number of subjects providing each rating varies (see Table 1). Attendees were told that “oral burn” referred to the sensation produced by putting an irritant into the mouth (e.g., chilis) and “oral pain” referred to an internally generated sensation (e.g., toothache, canker sore, biting the tongue).

8.2. General Labeled Magnitude Scale (gLMS)

The scale consisted of a horizontal line 11 cm long. The extreme left was labeled “no sensation” and the extreme right was labeled “strongest imaginable sensation of any kind.” Considering the line to be 100 units, the descriptors “barely detectable,” “weak,” “moderate,” “strong,” and “very strong” were spaced at 1.4, 6, 17, 35 and 53, respectively (the spacing determined by Green et al., 1993; Green, Dalton, Cowart, Rankin, & Higgins, 1996). To emphasize the general nature of the top of the scale, attendees were told to think of intense experiences from different modalities (e.g., staring at the sun, hearing a nearby jet plane take off, experiencing intense pain).

8.3. Subjects

When data are collected in a group setting, the experimenter has limited control over how well the subjects understand the instructions. For this reason, we required that data meet a criterion to be included. To be certain that subjects understood the instructions, data were only included if the subject rated “whisper” < “conversation” < “loudest sound experienced.” Of 406 subjects, 46 failed to meet this criterion and were excluded from this analysis. Of the 360 remaining, “loudest sound you have ever heard” and “brightest light you have ever seen” were tested for outliers (individuals falling > 2.5 S.D.s from the regression line of each rating plotted against PROP bitterness). Three subjects produced ratings of “brightest light experienced” that

Table 1
Correlations between remembered sensations and the bitterness of PROP paper

Remembered sensation	N	Mean intensity	Correlation with PROP bitterness Spearman Rho	P value
Brightness of a dimly lit restaurant	201	12.3	0.03	
Brightness of a well lit room	196	20.9	0.08	
Brightest light you have ever seen	283	68.8	0.02	
Loudness of a whisper	283	9.7	0.02	
Loudness of a conversation	283	24.9	0.11	
Loudest sound you have ever heard	283	73.9	0.12	<0.05
Warmth of warm bread in your mouth	279	29.3	0.12	
Strongest smell of a flower	176	44.7	0.14	
Sweetness of a coke	200	35.9	0.20	<0.01
Bitterness of celery	278	12.6	0.13	<0.05
Strongest saltiness experienced	282	54.1	0.12	<0.05
Strongest sweetness experienced	280	55.2	0.26	<0.0001
Strongest sourness experienced	281	59.1	0.24	<0.0001
Strongest bitterness experienced	283	56.1	0.18	<0.01
Strongest oral burn experienced	280	64.5	0.17	<0.01
Strongest oral pain experienced	280	62.4	0.10	
Maximum of oral burn and pain	280	72.8	0.12	<0.05
Candy	281	29.7	0.19	<0.01

were so low that they fell below the criterion and were excluded from this analysis.

Of the remaining 357 subjects, 283 had not previously tasted PROP; ratings from these subjects were used in the analyses. Of the 283, 156 were female, 113 were male and 14 did not indicate sex on the questionnaire.

8.4. Candy

Butterscotch discs purchased at Stop & Shop (Quincy, MA) were rated for sweetness.

8.5. PROP paper

PROP papers were made by soaking 3 cm diameter circular shaped pieces of filter paper (Whatman #1) in saturated pharmaceutical grade PROP (near boiling temperature). Papers were air dried and stored in small glassine envelopes. Each paper contains approximately 1.6 mg PROP. PROP crystallizes into the filter paper making it a convenient vehicle to deliver a measured amount of material into the mouth. Comparing the average perceived bitterness of the PROP papers with those of PROP solutions, the PROP paper falls between the perceived bitteresses of 0.001 and 0.0032 M PROP.

8.6. Procedure

Attendees first filled out the demographic information, then rated the foods/beverages for liking/disliking, next rated the sensations shown in Table 1 and rated the sweetness of the candy. The bitterness of the PROP paper was rated last to avoid context effects (Bartoshuk, 2000).

9. Results

Fig. 2 shows the most intense remembered sensations. ANOVA for the subjects rating all nine maximal sensations showed a highly significant effect of quality of sensation. Tukey HSD post-hoc tests (shown in Fig. 2) showed that the strongest flower smell was the least intense and sound the most intense of those sensory qualities tested for this group of subjects. Some subjects ($N=62$) rated two or more modalities at the same maximum. Of attendees who had only one sensation at the maximum ($N=221$) the sensations given the highest ratings were: 66 for sounds, 56 for oral pain, 38 for light, 34 for oral burn, 10 for bitter, 7 for sour, 6 for sweet and 4 for salty; no subject selected the strongest smell of a flower as the most intense of all of the sensations rated.

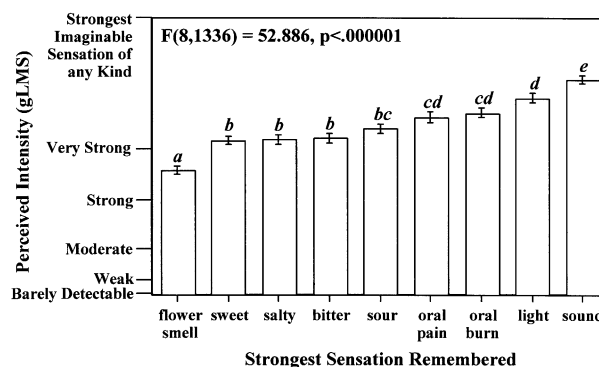


Fig. 2. Perceived Intensity (gLMS) $\pm$ SE of the most intense remembered sensations of nine qualitatively different sensations. Bars that do not share a letter are significantly different (at least $P < 0.05$) tested with a Tukey HSD post-hoc test (Keppel, 1991).

Table 1 shows the average rating for each of the remembered sensations as well as the Spearman rho correlations between these ratings and the bitterness of PROP paper. Spearman rho rather than Pearson r correlations were used because the PROP paper ratings are not normally distributed. Following the reasoning of Keppel (1991), we have not applied a Bonferroni correction since the number of planned correlations is small compared to the number of possible correlations. Note that the correlations are small. We cannot determine at this point whether they are small because they concern remembered rather than actual sensations (see Discussion) or whether they are small because of noise introduced by the group setting.

For 137 subjects, the bitterness of the PROP paper was rated higher than the “strongest bitterness experienced.” The correlation between PROP paper bitterness and the magnitude of the difference (strongest bitterness experienced–bitterness of PROP paper) was $\rho = 0.79$, $P < 0.0001$. For 36 subjects, the sweetness of the candy was rated higher than the “strongest sweetness experienced.” The bitterness of PROP paper was not correlated with the magnitude of the difference (strongest sweetness experienced–sweetness of candy) ($\rho = 0.04$, N.S.).

An examination of individuals at the extreme ends of the PROP distribution (bottom and top 10%) provides further information about the relations between oral burn, oral pain and PROP bitterness. The distributions for oral burn/oral pain of these two groups were significantly different (Chi-square = 6.72; $df = 1$; $P < 0.01$). On average, for subjects at the top of the PROP distribution (supertasters) the ratio of oral burn to oral pain was 1.63; for subjects at the bottom (nontasters), the ratio of oral burn to oral pain was 0.99 ($t = 2.639$, $P = 0.01$).

10. Discussion

10.1. Maximal perceived intensities are not the same across modalities

The results of the present study demonstrate that, contrary to the assumptions of range theory, maximum perceived intensities vary across modalities. Previous studies have shown that nontasters, medium tasters and supertasters live in different oral sensory worlds. The correlations of remembered sensations with PROP paper bitterness show that these differences can even be demonstrated with remembered sensations. Given the stability of spacing among intensity adjectives, this supports the position that adjective/adverb-labeled scales stretch and compress to fit the domain to which they are applied. Since the scales stretch and compress differently in groups with systematically different experiences, the

scales cannot be used to make valid comparisons across the groups.

The range of the maxima for various sensations in the real world is considerably larger than that reflected in Fig. 1. We selected “strongest smell of a flower” because we believed it would be obviously weaker than other items like “loudest sound you have ever heard.” However other sensory/hedonic experiences could have even weaker maxima. These might include sensations like those associated with ingestion (e.g., hunger, satiety, thirst) as well as some emotional experiences (e.g., annoyance, satisfaction).

10.2. Strongest remembered taste sensations correlate with perceived bitterness of PROP

Since none of the subjects in this study had previously tasted PROP, we can conclude that the differences in the taste worlds of nontasters, medium tasters and supertasters are sufficient to be reflected in subjects’ remembered taste sensations. But beyond this, these data are interesting in the context of remembered sensations in general. Of particular interest, about half of the subjects rated the PROP paper as more bitter than they had previously rated “strongest bitterness experienced;” the size of this difference was highly correlated with PROP status. Intense bitter tastes are not part of our ordinary experience and it is quite likely that many supertasters have not previously tasted anything as bitter as PROP tastes to them. Thus rating the PROP paper as more bitter than the “strongest bitterness experienced” could simply have been an accurate reflection of their experiences. Only 36 subjects rated the candy as more sweet than they had previously rated “strongest sweetness experienced;” the size of this difference was not correlated with PROP status.

Studies of remembered sensations suggest that they differ from actual sensations in systematic ways (e.g., see Algom, 1992, for a review). Thus rating the PROP paper as more bitter than “strongest bitter experienced” and/or rating the candy as sweeter than “strongest sweet experienced” could reflect faded memories of taste. Further, the differences across nontasters, medium tasters and supertasters for remembered taste sensations are smaller than the differences found when stimuli were actually tasted in previous studies. For example, saltiness (1.0 M NaCl) increased more across nontasters, medium tasters and supertasters in a study where the NaCl was actually tasted (Prutkin et al., 2000) than it did in the present study. This suggests the possibility that the fate of remembered taste sensations is different across nontasters, medium tasters and supertasters. This possibility deserves evaluation in an experiment designed to compare remembered and actual taste sensations in the same subjects. The properties of remembered sensations are important in the

context of scaling methodologies since the labels on scales usually refer to remembered sensations.

10.3. *The LMS and the gLMS*

The LMS was originally derived for oral sensations. Although we know that taste varies across nontasters, medium tasters and supertasters, oral burn/pain shows variation only for the tongue and not for other oral areas (e.g., floor of the mouth; Karrer et al., 1992). If the “strongest imaginable oral sensation” depended on those non-tongue areas, then the LMS could provide valid comparisons of oral sensations across nontasters, medium tasters and supertasters. The data presented here suggest that the burn perceived on the tongue tends to be more intense than oral pain for extreme supertasters. When one can establish that there are systematic differences in the absolute intensities experienced by specific groups, then the LMS cannot be used to provide valid comparisons across those groups. These data suggest that the gLMS rather than the LMS should be used for making comparisons across nontasters, medium tasters and supertasters of PROP.

10.4. *Invalid across-group comparisons*

We noted earlier that the issues raised here do not affect within-subject comparisons or across-group comparisons when the groups have been randomly selected. The invalid comparisons result when groups have experienced systematically different maxima in a given domain. For example, consider pain. Are there systematic differences in pain experiences between certain groups of individuals? Some reflection suggests this is certainly the case. Childbirth is said to be one of the most intense pains (Melzack, 1984). This suggests that the “worst pain you have experienced” is likely to reflect, on average, a less intense pain to men than to women who have borne children. Or consider food related sensations like hunger and satiety. Are these sensations blunted or exacerbated in individuals with eating disorders? In order to compare groups whose maximal sensations vary, on average, we must use a standard that is not related to the experience to be measured. Note that this means that the typical intensity scales (category, Likert, VAS) labeled in terms of minimal and maximal experiences of the domain to be measured will not provide valid across-group comparisons.

One of the issues that may have confused users of intensity labeled scales is the availability of validation studies. However, validation studies are aimed at ensuring that the measure at hand measures what the investigators want to measure. This can insure validity for within-subject comparisons but does not assure validity for across-subject/group comparisons. For example, in a validation of the VAS for mood (Nyenhuys,

Stern, Yamamoto, Luchetta, & Arruda, 1997), the VAS scores were correlated with a variety of other tests for mood. Significant correlations were taken as evidence of validity. But this can only establish similarity of relative positions on a variety of scales. This does not prove that the scales can then be used to compare absolute levels of moods across groups.

In some cases, investigators do not label a scale with the experience to be measured but rather simply use intensity descriptors (e.g., a line labeled nothing at one end and “extremely” at the other end; in other cases, investigators do not label the scale at all. In these cases, the subject must decide on the domain. Presumably, the context of the experiment provides this (e.g., a subject invited to participate in a taste study expects to rate tastes). Not specifying a domain does not solve any of the problems discussed here. Rather, it just adds more uncertainty about the nature of the subject’s context.

10.5. *An association between taste and hearing?*

Table 1 shows that the correlation between PROP paper bitterness and the loudest sound ever experienced was significant. An association between taste and hearing would be particularly important as we currently use sound as a standard in PROP studies. It is possible that such an association might result because certain kinds of damage to the ear could also damage the chorda tympani taste nerve (that nerve traverses the middle ear right behind the malleus bone on its way to the brain). A potential association between hearing and taste suggests the need for vigilance when considering a sensory standard. One of the ways in which we try to guard against this in laboratory studies is to administer a spatial taste test (Prutkin, Fast, Lucchina, Snyder, & Bartoshuk, 1999) to look for potential damage to the chorda tympani. A better standard than sound for PROP studies (i.e., a standard not at all related to taste), may well show that PROP effects are even larger than they appear to be in our recent studies.

10.6. *Advantages to improved scaling*

There may be many situations in which existing across-group comparisons will be revealed as valid. One easy way to test for group differences in sensations would be to assess those sensations (remembered or actual) with the gLMS or magnitude matching (using a sensory standard unrelated to the sensations of interest).

When scale usage proves to be invalid, correction of this will be to the advantage of the investigator; real group differences will be larger and some group differences previously obscured will be revealed. In a few cases, group differences erroneously thought to go in one direction will prove to go in the other. Correcting such instances will clearly be of value.

10.7. Combine the gLMS and magnitude matching with real and remembered sensations

The present data suggest a new psychophysical standard that should work for PROP studies and may well work in a variety of other settings as well. The “brightest light you have ever seen” did not correlate with PROP paper bitterness. Thus using that as a standard on a scale would permit assessment of differences across PROP groups. When questioned, not surprisingly, subjects tended to identify the sun as the brightest light ever seen. The brightness of the sun need not be perceived equally by all to make it a useful anchor for a sensory ruler; it need only be unrelated to the sensory maximum of interest.

The use of the gLMS and/or magnitude matching (with real or remembered sensations as standards) are not perfect solutions to comparing sensory and hedonic intensities across individuals/groups; however, they represent clear improvements over invalid comparisons. The best strategy we can recommend at present is to use a combination of methods (e.g., Chapo, Phillips et al., 2001). In an experiment using a ratio scale (i.e., gLMS, magnitude estimation), real or remembered standards can be included in the study and the data can be normalized to each of these standards (see, Bartoshuk, 2000, for an example of normalization). This provides converging operations that serve to insure that the conclusions of the study hold over a variety of assumptions.

Acknowledgements

Portions of these data were presented at the 2001 annual meetings of the Association for Chemoreception Sciences (Fast et al., 2001) and the Society for the Study of Ingestive behavior (Bartoshuk et al., 2001). Supported by NIH (DC00283 and AG18619) and NRICGP/USDA (9603745).

References

- Aitken, R. C. B. (1969). Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine*, 62, 989–993.
- Aitken, R. C. B., Ferres, H. M., & Gedy, J. L. (1963). Distraction from flashing lights. *Aerospace Medicine*, 34, 302–306.
- Algom, D. (1992). Memory psychophysics: an examination of its perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 441–513). New York: North-Holland.
- Bartoshuk, L. M. (1979). Bitter taste of saccharin: Related to the genetic ability to taste the bitter substance 6-n-propylthiouracil (PROP). *Science*, 205, 934–935.
- Bartoshuk, L. M. (1991). Sweetness: history, preference, and genetic variability. *Food Technology*, 45, 108–113.
- Bartoshuk, L. M. (1993a). The biological basis of food perception and acceptance. *Food Quality and Preference*, 4, 21–32.
- Bartoshuk, L. M. (2000). Comparing sensory experiences across individuals: recent psychophysical advances illuminate genetic variation in taste perception. *Chemical Senses*, 25, 447–460.
- Bartoshuk, L. M. (2002). Self reports and across-group comparisons: a way out of the box. *APS Observer*, 15, 7 26–28.
- Bartoshuk, L. M. (1993b). Genetic and pathological taste variation. In D. Chadwick, J. Marsh, & J. Goode (Eds.), *What can we learn from animal models and human disease? (Ciba Foundation Symposium 179)* (pp. 251–267). New York: John Wiley and Sons.
- Bartoshuk, L. M., Caseria, D., Catalanotto, F., Dabril, G., Duffy, V. B., Lucchina, L. A., Nadoolman, W., Sasaki, C., Snyder, D. J., & Wolfe, J. (1996). Do taste-trigeminal interactions play a role in oral pain? *Chemical Senses*, 21, 578.
- Bartoshuk, L. M., Conner, E., Grubin, D., Karrer, T., Kochenbach, K., Palco, M., Snow, D., Pelchat, M., & Danowski, S. (1993). PROP supertasters and the perception of ethyl alcohol. *Chemical Senses*, 18, 526–527.
- Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B., Kveton, J., Lucchina, L. A., Prutkin, J. M., Snyder, D. J., & Tie, K. (1999). Sensory variability, food preferences and BMI in non-, medium and supertasters. *Appetite*, 33, 228–229.
- Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B. G., & Snyder, D. J. (2001). The General Labeled Magnitude Scale provides valid measures of genetic variation in taste and may be a universal psychophysical ruler. *Appetite*, 37, 126.
- Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B. G., & Snyder, D. J. (in press). Hormones, age, genes and pathology: How do we assess variation in sensation and preference? In *Food selection, from genes to culture*. Paris, France: Danone Institute.
- Bartoshuk, L. M., Duffy, V. B., Lucchina, L. A., Prutkin, J., & Fast, K. (1998). PROP (6-n-propylthiouracil) supertasters and the saltiness of NaCl. In C. Murphy (Ed.), *Olfaction and taste XII (Vol. 855)* (pp. 793–796). New York: New York Academy of Sciences.
- Bartoshuk, L. M., Duffy, V. B., & Miller, I. J. (1994). PTC/PROP tasting: anatomy, psychophysics, and sex effects. *Physiology and Behavior*, 56, 1165–1171.
- Bartoshuk, L. M., Duffy, V. B., Reed, D., & Williams, A. (1996). Supertasting, earaches, and head injury: genetics and pathology alter our taste worlds. *Neuroscience and Biobehavioral Reviews*, 20, 79–87.
- Bartoshuk, L. M., Fast, K., Duffy, V. B., Prutkin, J. M., Snyder, D. J., & Green, B. G. (2000). Magnitude matching and a modified LMS produce valid sensory comparisons for PROP studies. *Appetite*, 35, 277.
- Bartoshuk, L. M., Fast, K., Karrer, T. A., Marino, S., Price, R. A., & Reed, D. A. (1992). PROP supertasters and the perception of sweetness and bitterness. *Chemical Senses*, 17, 594.
- Bartoshuk, L. M., Rifkin, B., Marks, L. E., & Hooper, J. E. (1988). Bitterness of KCl and benzoate: related to genetic status for sensitivity to PTC/PROP. *Chemical Senses*, 13, 517–528.
- Berry, H., & Huskisson, E. C. (1972). Treatment of rheumatoid arthritis. *Clinical Trials Journal*, 9, 13–14.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgements. *Journal of Personality and Social Psychology*, 66, 5–20.
- Birnbaum, M. H. (1999). How to show that 9 > 221: collect judgements in a between-subjects design. *Psychological Methods*, 4, 243–249.
- Blakeslee, A. F. (1932). Genetics of sensory thresholds: taste for phenyl thio carbamide. *Proceedings of the National Academy of Sciences*, 18, 120–130.
- Borg, G. (1961). Interindividual scaling and perception of muscular force. *Kungl. Fysiografiska Sällskapet i Lund Förhandlingar*, 31, 117–125.
- Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H. G. Geissler, & P. Petzold (Eds.), *Psychophysical judgment and the process of perception* (pp. 25–34). Berlin: VEB Deutscher Verlag der Wissenschaften.
- Borg, G. (1992). A “fixed star” for interprocess comparisons. In G. Borg, & G. Neely (Eds.), *Fechner Day 92: Proceedings of the eighth annual meeting of the international society of psychophysics* (pp. 41–45). Stockholm, Sweden: Stockholm University.

- Borg, G. (1994). Psychophysical scaling: an overview. In J. Boivie, P. Hansson, & U. Lindblom (Eds.), *Touch, temperature, and pain in health and disease: mechanisms and assessments* (pp. 27–39). Seattle, WA: IASP Press.
- Champney, H. (1941). The measurement of parent behavior. *Child Development*, 12, 131–166.
- Chapo, A. K., Bartoshuk, L. M., Peterson, J. M., Phillips, M. N., & Duffy, V. B. (2001). Salt intensity and behaviors: associations with bitterness of 6-n-propylthiouracil (PROP). *Chemical Senses*, 26, 1072.
- Chapo, A. K., Phillips, M. N., Ilich, J. Z., & Duffy, V. B. (2001). Sodium chloride (NaCl) saltiness: are older females more responsive. *The Gerontologist*, 41, 83.
- Clarke, P. R. F., & Spear, F. G. (1964). Reliability and sensitivity in self-assessment of well-being. *Bulletin of the British Psychological Society*, 17, 18 A.
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66, 27–44.
- Duffy, V. B., & Bartoshuk, L. M. (1996). Sensory factors in feeding. In E. D. Capaldi (Ed.), *Why we eat what we eat: the psychology of eating* (pp. 145–171). Washington, DC: American Psychological Association.
- Duffy, V. B., Bartoshuk, L. M., Lucchina, L. A., Snyder, D. J., & Tym, A. (1996). Supertasters of PROP (6-n-propylthiouracil) rate the highest creaminess to high-fat milk products. *Chemical Senses*, 21, 598.
- Duffy, V. B., Phillips, M. N., Peterson, J. M., & Bartoshuk, L. M. (2001). Bitterness of 6-n-propylthiouracil (PROP) associates with bitter sensations and intake of vegetables. *Appetite*, 37, 137–138.
- Duffy, V. B., Fast, K., Cohen, Z., Chodos, E., & Bartoshuk, L. M. (1999). Genetic taste status associates with fat food acceptance and body mass index in adults. *Chemical Senses*, 24, 545–546.
- Duffy, V. B., & Peterson, J. M. (2000). Genetic variation in taste: associations with alcohol sensation and intake. *Chemical Senses*, 25, 638.
- Fast, K., Duffy, V. B., & Bartoshuk, L. M. (in press). New psychophysical insights in evaluating genetic variation in taste. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste and cognition*.
- Fast, K., Green, B. G., Snyder, D. J., & Bartoshuk, L. M. (2001). Remembered intensities of taste and oral burn correlate with PROP bitterness. *Chemical Senses*, 26, 1069.
- Fernberger, S. W. (1932). A preliminary study of taste deficiency. *American Journal of Psychology*, 44, 322–326.
- Finger, T. E., Nelson, G. M., Bryant, B., & Moore, P. A. (1994). Intragummal and perigemmal fibers in taste buds: Immunocytochemistry and differential sensitivity to capsaicin. *Neuroscience Abstracts*, 402, 12.
- Fischer, R. (1967). Genetics and gustatory chemoreception in man and other primates. In M. R. Kare, & O. Maller (Eds.), *The chemical senses and nutrition* (pp. 621–681). Baltimore: John Hopkins Press.
- Fox, A. L. (1931). Six in ten “tasteblind” to bitter chemical. *Science News Letter*, 9, 249.
- Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology*, 14, 83–102.
- Gent, J. F., & Bartoshuk, L. M. (1983). Sweetness of sucrose, neohesperidin dihydrochalcone, and saccharin is related to genetic ability to taste the bitter substance 6-n-propylthiouracil. *Chemical Senses*, 7, 265–272.
- Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology*, 39, 169–200.
- Gracely, R. H., McGrath, P., & Dubner, R. (1978). Ratio scales of sensory and affective verbal pain descriptors. *Pain*, 5, 5–18.
- Green, B. G., Shaffer, G. S., & Gilmore, M. M. (1993). A semantically-labeled magnitude scale of oral sensation with apparent ratio properties. *Chemical Senses*, 18, 683–702.
- Green, B. G., Dalton, P., Cowart, B., Rankin, K., & Higgins, J. (1996). Evaluating the labeled magnitude scale for measuring sensations of taste and smell. *Chemical Senses*, 21, 323–334.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hakel, M. D. (1968). How often is often?. *American Psychologist*, 23, 533–534.
- Hall, M. J., Bartoshuk, L. M., Cain, W. S., & Stevens, J. C. (1975). PTC taste blindness and the taste of caffeine. *Nature*, 253, 442–443.
- Harris, H., & Kalmus, H. (1949). The measurement of taste sensitivity to phenylthiourea (P.T.C.). *Annals of Eugenics*, 15, 24–31.
- Hayes, S., & Patterson, D. G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, 18, 98–99.
- Heft, M. W., & Parker, S. R. (1984). An experimental basis for revising the graphic rating scale for pain. *Pain*, 19, 153–161.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Hetherington, M. M., & Rolls, B. J. (1987). Methods of investigating human eating behavior. In F. M. Toates, & N. E. Rowland (Eds.), *Feeding and drinking* (pp. 77–109). New York: Elsevier Science (Biomedical Division).
- Jones, L. V., Peryam, D. R., & Thurstone, L. L. (1955). Development of a scale for measuring soldier's food preferences. *Food Research*, 20, 512–520.
- Kalmus, H. (1958). Improvements in the classification of the taster genotypes. *Annals of Human Genetics*, 22, 222–230.
- Karrer, T., & Bartoshuk, L. (1991). Capsaicin desensitization and recovery on the human tongue. *Physiology and Behavior*, 49, 757–764.
- Karrer, T., Bartoshuk, L. M., Conner, E., Fehrenbaker, S., Grubin, D., & Snow, D. (1992). PROP status and its relationship to the perceived burn intensity of capsaicin at different tongue loci. *Chemical Senses*, 17, 649.
- Kast, E. C. (1962). The measurement of pain: a new approach to an old problem. *The Journal of New Drugs*, 2, 344–351.
- Keppel, G. (1991). *Design and analysis: a researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Ko, C. W., Hoffman, H. J., Lucchina, L. A., Snyder, D. J., Weiffenbach, J. M., & Bartoshuk, L. M. (2000). Differential perceptions of intensity for the four basic taste qualities in PROP supertasters versus nontasters. *Chemical Senses*, 25, 639–640.
- Kong, A., Barnett, G. O., Mosteller, F., & Youtz, C. (1986). How medical professionals evaluate expressions of probability. *The New England Journal of Medicine*, 315, 740–744.
- Lasagna, L. (1960). The clinical measurement of pain. *Annals of the New York Academy of Sciences*, 86, 28–37.
- Lawless, H. T. (1980). A comparison of different methods used to assess sensitivity to the taste of phenylthiocarbamide (PTC). *Chemical Senses*, 5, 247–256.
- Lawless, H. T., Horne, J., & Spiers, W. (2000). Contrast and range effects for category, magnitude and labeled magnitude scales in judgements of sweetness intensity. *Chemical Senses*, 25, 85–92.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5–55.
- Lucchina, L. A., Curtis, O. F., Putnam, P., Drewnowski, A., Prutkin, J. M., & Bartoshuk, L. M. (1998). Psychophysical measurement of 6-n-propylthiouracil (PROP) taste perception. In C. Murphy (Ed.), *Olfaction and taste XII* (pp. 817–822). New York: New York Academy of Sciences.
- Mapes, R. E. A. (1979). Verbal and numerical estimates of probability in therapeutic contexts. *Social Science & Medicine*, 13A, 277–282.
- Marino, S., Bartoshuk, L. M., Monaco, J., Anliker, J. A., Reed, D., & Desnoyers, S. (1991). PTC/PROP and the tastes of milk products. *Chemical Senses*, 16, 551.
- Marks, L. E., & Stevens, J. C. (1980). Measuring sensation in the aged. In L. W. Poon (Ed.), *Aging in the 1980's: psychological issues* (pp. 592–598). Washington: American Psychological Association.
- Marks, L. E., Stevens, J. C., Bartoshuk, L. M., Gent, J. G., Rifkin, B., & Stone, V. K. (1988). Magnitude matching: the measurement of taste and smell. *Chemical Senses*, 13, 63–87.

- Melzack, R. (1984). The myth of painless childbirth. *Pain*, 19, 321–337.
- Michell, J. (1999). *Measurement in psychology*. Cambridge, UK: Cambridge University Press.
- Miller, I. J., & Reedy, F. E. (1990). Variations in human taste bud density and taste intensity perception. *Physiology and Behavior*, 47, 1213–1219.
- Moskowitz, H. R. (1977). Magnitude estimation: Notes on what, how, when, and why to use it. *Journal of Food Quality*, 1, 195–228.
- Moskowitz, H. R., & Chandler, J. W. (1977). New uses of magnitude estimation. In G. G. Birch, J. G. Brennan, & K. J. Parker (Eds.), *Sensory properties of foods* (pp. 189–211). London: Applied Science.
- Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, 5, 2–34.
- Nagy, J. I., Goedert, M., Hunt, S. P., & Bond, A. (1982). The nature of the substance P-containing nerve fibers in taste papillae of the rat tongue. *Neuroscience*, 7, 3137–3151.
- Narens, L., & Luce, R. D. (1983). How we may have been misled into believing in the interpersonal comparability of utility. *Theory and Decision*, 15, 247–260.
- Nyenhuis, D. L., Stern, R. A., Yamamoto, C., Luchetta, T., & Arruda, J. E. (1997). Standardization and validation of the visual analogue mood scales. *The Clinical Neuropsychologist*, 11, 407–415.
- Parducci, A. (1965). Category judgement: a range-frequency model. *Psychological Review*, 72, 407–418.
- Patterson, D. G. (1923). The Scott company graphic rating scale. *The Journal of Personnel Research*, 1, 361–376.
- Peck, R. E. (1966). The application of thymometry to the measurement of anxiety. *International Journal of Neuropsychiatry*, 2, 337–341.
- Peck, R. E. (1967). A precise technique for the measurement of pain. *Headache*, 6, 189–194.
- Peryam, D. R., & Girardot, N. F. (1952). Advanced taste test method. *Food Engineering*, 24, 58–61 194.
- Peterson, J. M., Bartoshuk, L. M., & Duffy, V. B. (1999). Intensity and preference for sweetness influenced by genetic taste variation. *Journal of the American Dietetic Association, Suppl.*, 99, A–28.
- Peterson, J. M., & Duffy, V. B. (2000a). Acceptance of sweetness: associations with genetic variation in taste and sex. *Chemical Senses*, 25, 638–639.
- Peterson, J. M., & Duffy, V. B. (2000b). Genetic variation in taste: associations with sweetness intensity, sweet liking and sweet food acceptance. *Chemical Senses*, 25, 638–639.
- Phillips, M. N., Bartoshuk, L. M., Peterson, J. M., & Duffy, V. B. (2001). Genetic variation in taste: Associations with creamy sensations, preference for and intake of high fat foods. *Chemical Senses*, 26, 1040–1041.
- Prescott, J., Johnstone, V., & Munro, P. (2001). Discriminability of fat content as a function of PROP sensitivity. *Chemical Senses*, 26, 800.
- Prescott, J., & Swain-Campbell, N. (2000). Responses to repeated oral irritation by capsaicin, cinnamaldehyde and ethanol in PROP tasters and non-tasters. *Chemical Senses*, 25, 239–246.
- Price, D. D., McGrath, P. A., Rafii, A., & Buckingham, B. (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain*, 17, 45–56.
- Prutkin, J., Duffy, V. B., Etter, L., Fast, K., Gardner, E., Lucchina, L. A., Snyder, D. J., Tie, K., Weiffenbach, J., & Bartoshuk, L. M. (2000). Genetic variation and inferences about perceived taste intensity in mice and men. *Physiology and Behavior*, 69, 161–173.
- Prutkin, J. M., Fast, K., Lucchina, L. A., Snyder, D. J., & Bartoshuk, L. M. (1999). Spatial taste testing and genetic taste variation. *Chemical Senses*, 24, 604.
- Prutkin, J. M., Fast, K., Lucchina, L. A., & Bartoshuk, L. M. (1998). *Prop (6-n-propylthiouracil) genetics and trigeminal innervation of fungiform papillae*. Paper presented at the XXth Meeting of the Association for Chemoreception Sciences, Sarasota, FL.
- Reed, D. R., Bartoshuk, L. M., Duffy, V., Marino, S., & Price, R. A. (1995). Propylthiouracil tasting: determination of underlying threshold distributions using maximum likelihood. *Chemical Senses*, 20, 529–533.
- Reed, D. R., Li, X., Chen, Z., Mascioli, K., Bachmanov, A. A., Beauchamp, G. K., Tordoff, G. K., Max, M., Margolskee, R. F., Bartoshuk, L. M., & Duffy, V. B. (2002). Alleles of the putative taste receptor gene TAS1R3 may influence the pleasantness of sucrose and aspartame in human subjects. *Obesity Research*, 10, 67.
- Schutz, H. G., & Cardello, A. V. (2001). A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *Journal of Sensory Studies*, 16, 117–159.
- Silver, W. L., & Finger, T. E. (1991). The trigeminal system. In T. V. Getchell, R. L. Doty, L. M. Bartoshuk, & J. B. Snow (Eds.), *Smell and taste in health and disease* (pp. 97–108). New York: Raven Press.
- Silverstone, T., & Stunkard, A. J. (1968). The anorectic effect of dex-amphetamine sulphate. *British J. Pharmacol. Chemother.*, 33, 513–522.
- Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. *The Quarterly Journal of Speech*, 30, 328–330.
- Snyder, D. J., Duffy, V. B., Fast, K., Hoffman, H. J., Ko, C. W., Weiffenbach, J. M., & Bartoshuk, L. M. (2001). Food preferences vary with age and sex: a new analysis using the general Labeled Magnitude Scale. *Chemical Senses*, 26, 1050.
- Snyder, D. J., Duffy, V. B., Fast, K., Weiffenbach, J. M., & Bartoshuk, L. M. (2001). PROP genetics interact with age and sex to influence food preferences. *Appetite*, 37, 164.
- Snyder, D. J., Lucchina, L. A., Duffy, V. B., & Bartoshuk, L. M. (1996). Magnitude matching adds power to the labeled magnitude scale. *Chemical Senses*, 21, 673.
- Snyder, L. H. (1931). Inherited taste deficiency. *Science*, 74, 151–152.
- Stevens, J. C. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 57, 201–209.
- Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, 54, 407–411.
- Stevens, J. C., & Marks, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception and Psychophysics*, 27, 379–389.
- Stevens, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 69, 1–25.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64, 153–181.
- Stevens, S. S. (1958). Adaptation-level vs the relativity of judgment. *The American Journal of Psychology*, 4, 633–646.
- Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 54, 377–411.
- Teghtsoonian, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' law. *American Journal of Psychology*, 86, 3–27.
- Tepper, B. J., & Nurse, R. J. (1997). Fat perception is related to PROP taster status. *Physiology and Behavior*, 61, 949–954.
- Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *The Knowledge Engineering Review*, 10, 43–62.
- Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability words. *Journal of Experimental Psychology: General*, 115, 348–365.
- Weiffenbach, J. M., Duffy, V. B., Fast, K., Cohen, Z. D., & Bartoshuk, L. M. (2000). Bitter-sweet age sex and PROP (6-n-propylthiouracil) effects: a role for menopause? *Chemical Senses*, 25, 639.

- Whitehead, M. C., Beeman, C. S., & Kinsella, B. A. (1985). Distribution of taste and general sensory nerve endings in fungiform papillae of the hamster. *American Journal of Anatomy*, 173, 185–201.
- Zoeke, B., & Sarris, V. (1983). A comparison of 'frame of reference' paradigms in human and animal psychophysics. In H.-G. Geissler, J. F. J. M. Buffant, E. L. J. Leeuwenburg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 283–317). Berlin: VEB Deutscher Verlag der Wissenschaften.