The labeled dissimilarity scale: A metric of perceptual dissimilarity

DANIEL B. KURTZ, THERESA L. WHITE, and MEGHAN HAYES
SUNY Health Science Center, Syracuse, New York

Fundamental to the concept of *psychological distance* is the idea that confusability allows discovery of the perceptual relationships between objects, which provides understanding of the underlying principles that govern the functioning of a system. Thus, judgments of dissimilarity (conceptually proportional to the inverse of confusability) may provide insight into the elusive underlying quality-coding mechanisms in that sensory system. In the present experiments, a labeled dissimilarity scale (LDS) that reflects the magnitude of odorant dissimilarity was developed in a fashion similar to that reported by Green (Green, Shaffer, & Gilmore, 1993). This scale was produced by rating the perceptual intensity implied by adverbs describing different levels of dissimilarity, and then attaching those descriptors to appropriate locations on a numerical scale. The usefulness of the scale was demonstrated by its ability to produce visual color space with ratings of dissimilarity of Munsell color chips. The stability and reliability of the LDS was evaluated by comparing it with the traditional scaling technique of magnitude estimation (ME). It was found that the scales produced similar ratings of odorant dissimilarity and showed a similar susceptibility to the effects of contrast convergence. However, the coefficients of variation of dissimilarities rated with ME were much higher than those produced with the LDS. The subjects also dealt with the LDS without the anxiety that usually accompanies first-time users of ME. The LDS provided stable ratings of odorant dissimilarity and preserved the inferred ratio scale properties of ME.

The concept of *psychological distance*, as embodied by dissimilarity between stimuli, has proven useful in the understanding of many perceptual and cognitive processes. Investigations of dissimilarity take a number of forms, including errors in identification that are often summarized in confusion matrices (actually a measure of confusability), measures of discrimination, and direct scaling of perceptual similarity/dissimilarity. Regardless of the form, the premise of investigations of psychological distance is that dissimilarity allows for the discovery of the psychological or perceptual relationships between objects, thus suggesting the principles that govern the functioning of a system.

A prime example of the utility of the concept of psychological distance can be found through the evaluation of color vision in normal and colorblind people. It is a simple task to ask people with normal color vision to rate the degree of dissimilarity between the hues of various

wavelengths of light. The ratings clarify relationships between colors and result in the familiar color wheel (Shepard, 1962). The color wheel is a type of visual color object space that represents the perceptual properties of different wavelengths of light. The wavelengths that are close to each other in this space are similar in hue, whereas wave lengths that are distant in space are dissimilar in hue. In a similar fashion, visual object space can be explored in color-vision-impaired individuals. For example, ratings of colors will be altered in people with dichromatic vision such that all wavelengths appear to be mapped to one short wavelength (blue) and one long wavelength (yellow). That is, many wavelengths are mapped to two narrow color perceptions. Thus, the underlying relationships that are revealed through knowledge of the specific similarities between different wavelengths allow for the differential diagnosis of color deficiencies (Farnsworth, 1943).

In much the same way as in color vision, relationships between odorant qualities have been the basis for investigating the underlying principles in olfactory quality coding. Many studies (e.g., Bienfang, 1941; Davis, 1979; Kemp, Ishii, & O'Mahony, 1996; Schiffman, Robinson, & Erickson, 1977; Woskow, 1966; Yoshida, 1964) have used odor similarities as the basis for investigation into the structure and dimensionality of odorant object space. Generally, the primary axis, or dimension, seems to be related to *hedonic tone*—the degree to which an odorant is pleasant or unpleasant (Schiffman, 1974). The identification of the remaining dimensions, generally being related to the specific set of odorants under investigation, is largely idiosyncratic. Unfortunately, no simple structure

The authors express their gratitude to Maxwell Mozell and Lawrence Marks, both of whom offered excellent advice on an earlier draft of this paper. The authors also thank the reviewers of this article, particularly Barry Green, for their helpful comments on earlier drafts of the manuscript. Experiment 2 was presented at the 1997 International Symposium of Olfaction and Taste, with the abstract of that presentation published in *Chemical Senses*, Volume 22 (6). Preparation of this article was supported in part by NIH Grant 9-PO1 DC00220. Requests for reprints should be sent to D. B. Kurtz, SUNY Upstate Medical University, Department of Neuroscience and Physiology, 750 East Adams St., Syracuse, NY 13210 (e-mail: kurtzd@vax.cs.hscsyr.edu).

—Accepted by previous editor, Myron L. Braunstein

Table 1
Verbal Stimuli Used in Experiment 1

absolutely	extremely	negligibly
appreciably	fairly	partially
barely	hardly	perceptibly
clearly	marginally	radically
completely	markedly	significantly
considerably	mildly	slightly
discernably	moderately	somewhat
entirely	modestly	substantially

has emerged to describe the systematic relationships between odors. It is possible that no structure exists or that if one does exist, it has a high dimensionality. Indeed, it may be that odorant object space has a dimensionality that is one less than the vast number of odorant receptor proteins implied in the work of Buck and Axel (1991; Buck, 1993).

Although the overall structure of odorant object space remains unknown, differences in odor quality perception may provide clues to the coding of specific odor qualities. O'Connell and Stevens (O'Connell, Stevens, Akers, Coppola, & Grant, 1989; D. A. Stevens & O'Connell, 1996) examined similarities between the quality of specific odorants in their investigations of specific anosmias, conditions in which there is an insensitivity to a particular odorous stimulus (Amoore, 1977). Differences in odorant quality perception between people suggested that odorant molecules interact with more than one perceptual channel or receptor process (D. A. Stevens & O'Connell, 1991).

The discussion above suggests that judgments of dissimilarity, or similarity, between odorants could be useful in the evaluation of olfactory function, as well as dysfunction. Furthermore, judgments regarding the differences between odorants may provide insight into the elusive underlying olfactory quality-coding mechanisms. However, the question remains as to the type of dissimilarity rating scale that will most effectively and easily elicit these types of judgments from subjects.

One method of obtaining dissimilarity estimates from subjects is category scaling (Kuznicki, 1980). In this technique, the subject is presented with stimuli and asked to describe the difference between them as one of a specified number of categories. These categories may be verbal (such as *low*, *medium*, or *high*) or numeric (such as 1, 2, or 3). One difficulty that can arise with category scales, however, is that often each of the categories is treated as arising from sensations that are equidistant from each other, an assumption that is not necessarily true. Furthermore, a strong tendency exists for observers to assign categories to stimuli such that all categories are used equally often, thus distorting the shape of the psychophysical function (Gescheider, 1976, pp. 108–109).

In the scaling technique of magnitude estimation (ME; J. C. Stevens, 1971), subjects assign a number that represents the degree of difference between the two stimuli. Although it has been proposed that subjects are able to perform this task through some absolute internal scale (e.g.,

Zwislocki & Goodman, 1980), naive subjects find ME difficult and generally need to be trained in order to execute the task without idiosyncratic number usage (Green, Shaffer, & Gilmore, 1993).

Borg (1982) developed an alternative scaling method to ME called the *category ratio* (CR) scale. In the course of attempting to measure the perceived level of physical exertion, Borg postulated that when people are at their maximum level of work they have essentially the same perceived exertion. To compare sensations across subjects, the CR was bounded by the semantic category *maximum level of exertion* at the top, and *nothing at all* at the bottom, both of which were theoretically common across subjects. In addition, Borg anchored a number of locations on the scale with verbal labels, exploiting the multiplicative qualities of adverbs (Cliff, 1959). This scale approximated a logarithmic scale in that specific descriptors were placed at specific locations on a 10-point scale (Borg, 1982). In a later modification of this scale, verbal labels were arranged along a logarithmic continuum that, because of their numeric values, caused the verbal labels to lie at approximately equal intervals (Marks, Borg, & Ljunggren, 1983).

Recently, a *labeled magnitude scale* (LMS) was developed (Green et al., 1993) for the study of the magnitude of oral sensations, inspired by the CR scale developed by Borg (1982). The LMS combined the ratio properties of ME with the directness and convenience of a category scale. Similar to the previous scales, the LMS was fashioned with the unequal (nearly logarithmic) positioning of the semantic categories of intensity, as well as the bounding of the scale by *strongest imaginable* and *barely detectable* at each end. Unique to the LMS was that the numerical locations of the labels were determined directly by ME of the intensity of the verbal descriptors relative to other oral sensations. A comparison of this scale with ME indicated that both methods produced similar psychophysical functions, and thus demonstrated that the LMS produced perceptual ratings with the same ratio scale properties as those for ME without the training that was necessary for subjects to use that scale (Green et al., 1993; Zheng, Hoffman, Lucchina, Bartoshuk, & Weiffenbach, 1998).

In the present study, the design of an LDS for rating odorant dissimilarity conceptually comprised two parts: (1) selecting labels that described various levels of difference and, (2) positioning the selected labels in the appropriate locations on a scale bounded by maximal and minimal sensations (Borg, 1982).

In a preliminary experiment, 20 subjects ranked the 24 adverbs listed in Table 1 from lowest to highest in order of degree of dissimilarity.

A frequency analysis of the ranks of the adverbs was performed, with ties treated as equal to the mean of the ranks involved [i.e., three items tied with a rank of 4 would each be scored as $(4 + 5 + 6)/3$, or 5].

Adverbs were selected for inclusion in Experiment 1 on the basis of the following criteria: (1) preference was

Table 2
Average Ranks of Adverbs Generated in Pilot Data

Adverb	Average	SD
barely	1.73	1.12
hardly	3.28	1.62
mildly	6.35	2.41
partially	8.88	2.46
moderately	10.83	2.11
considerably	15.35	2.62
substantially	18.48	1.59
extremely	21.40	1.13
absolutely	23.13	1.23

given to adverbs with the smallest range of ranks and (2) preference was given to adverbs that did not show idiosyncratic scaling. That is, the number of adverbs with unique rankings by a single subject was minimized. These criteria were developed to select adverbs with meanings for which there was a consensus across subjects. As a side effect of these criteria, the chosen adverbs were greater than 0.90 standard deviations away from the mean of other adverbs. These adverbs are listed from lowest to highest in Table 2, along with the means and standard deviations of their ranks.

EXPERIMENT 1 Relationships Between Words

Having selected the adverbs that the subjects generally placed in the same order, the next step in developing an LDS was to attach numbers to the magnitude of dissimilarity that these adverbs represented. The appropriate placement of labels along the scale was critical in providing semantic information about the degree of difference indicated by the scale. Green et al. (1993) accomplished this aspect of scale construction by asking the subjects to provide magnitude estimates of the perceptual intensity implied by a variety of verbal descriptors. In the present experiment, we adopted a similar approach. The experiment represented an opportunity to both verify the rank orders obtained in the pilot experiment and refine the descriptors attached to the dissimilarity scale.

Method

Subjects. Twenty subjects (10 men, 10 women) participated in the experiment (3 subjects had participated in the preliminary experiment). The subjects were selected so that half of them (5 men, 5 women) were above 50 years of age (range = 52–69, $M = 58.5$, $\sigma^2 = 6.85$), and the other half, less than 50 years old (range = 21–46, $M = 32.7$, $\sigma^2 = 8.05$).

Stimuli. The nine adverbs (selected in the pilot experiment) that are listed in Table 2 were printed on separate pieces of paper measuring approximately 1×5 cm. Each adverb was followed by the word *different*, in the form “____ Different.”

Procedure. A wooden meterstick that was labeled “Not Different” at 0 cm and “Most Different Imaginable” at 100 cm was given to the subjects along with the stimuli. The subjects were verbally instructed to perform a two-step task. The first step was to organize the nine descriptors in order of increasing magnitude of difference, according to their own semantic interpretation. Once ordered, the descriptors were to be placed along the meterstick such that the

spacing reflected their relationship to one another. The subjects were told that if a word represented twice as much difference as that for another word it should be placed twice as far from “Not Different” but within the context set by the two fixed labels. (Note: This procedure was employed to allow the subjects maximum freedom in ratio production. In a pilot study reported below, when judging perceptual intensities with ME, the subjects tended to assign whole numbers to the verbal descriptors. Although this resulted in positioning that was similar to the position of many descriptors employed in Borg’s 1982 scale, it did not result in the fine-tuning of descriptor magnitude that was desired.)

Results

The position of each descriptor, to the closest centimeter marking, was recorded. The mean rating and standard deviation for each descriptor are shown in Table 3. The data were evaluated in a $2 \times 2 \times 9$ analysis of variance (ANOVA) with the factors being age, gender, and descriptors. In this comparison, little difference was seen between the positioning of descriptors by older and younger groups [$F(1,16) = 0.0093$, $p > .25$], between women and men [$F(1,16) = 0.5633$, $p > .25$], or in the interaction of the two factors [$F(1,16) = 0.3623$, $p > .25$]. As expected, the effect of descriptors was highly significant [$F(8,128) = 410.02$, $p < .001$], but there were no interactions between descriptors, age, or gender. With the exception of *partially*, which had a large standard deviation, each of the words could have been chosen to be included in the final scale.

Choice of Scale Descriptors

Descriptors were chosen in order to avoid standard deviation overlap. The descriptors meeting this criterion were placed along the scale at the following locations: *barely different* at 5.1, *mildly different* at 18.5, *moderately different* at 42.8, *considerably different* at 65.6, and *extremely different* at 87.6. Because some of the subjects in this experiment rated *absolutely different* above 100 (labeled as “Most Different Imaginable”) the scale was extended to 106 to allow for ratings of exceptional dissimilarity. Because of this ambiguity, subsequent scales did not have either *most different imaginable* or *absolutely different* attached near the top of the scale. In a similar fashion, Borg (1982) moved “Maximal” from the top of the scale to a position off the end of the scale. Numeric labels were placed every 10 units, beginning at 0, and marks were placed every 2 for accurate scoring. The resultant LDS is shown in Figure 1.

Table 3
Average Ratings of Adverbs Generated in Experiment 1

Adverb	Average	SD
barely	5.10	3.39
hardly	9.55	6.04
mildly	18.50	6.58
partially	30.25	12.17
moderately	42.80	8.37
considerably	65.60	9.63
substantially	70.35	8.37
extremely	87.60	5.59
absolutely	96.20	3.71

Figure 1. The labeled dissimilarity scale for odor dissimilarity.

Discussion

The selection of adverbs that describe levels of difference was a critical part of the construction of the LDS. Agreement upon the level of difference implied by each adverb was critical in creating a scale that would be of widespread utility. In the present experiment, the ordering of the adverbs confirmed the results of the pilot experiment. The subjects in both the pilot and the present experiment agreed upon the order of the rankings of the five descriptors chosen for the final scale and, therefore, also agreed to some degree on the relative level of difference suggested by the adverbs. However, the primary focus of the present experiment was to attach numeric values to the adverbs, thus making more specific the general agreement on the level of difference implied by each word.

It is interesting that the words were spaced at approximately equal intervals on the LDS. It is possible that the subjects ignored the instruction to be mindful of ratio relationships between the words, and spaced the words so

that each word represented a fixed portion of the meterstick (e.g., a category scale). To examine this possibility, the values obtained with the meterstick were compared with the values obtained through ME in an earlier pilot experiment. In the pilot experiment, the subjects were first trained in ME with projected line lengths and then, in a similar fashion, were asked to judge the subjective magnitude implied by each of the following words: *indistinguishable*, *just different*, *slightly different*, *moderately different*, *mildly different*, *substantially different*, and *extremely different*. The following four descriptors from the pilot experiment that overlapped with descriptors from the present experiment (assuming that *not different* from the present experiment and *indistinguishable* from the pilot experiment represent the same perceptual intensity) are: *not different*, *mildly different*, *moderately different*, and *extremely different*. The ME values were normalized in such a way that the average of all four ratings equaled the average of all four meterstick ratings for the same words. As is shown in Figure 2, there is a slight negative curvature to this relationship, indicating a bit of a ceiling effect when the words were placed on the meterstick. However, most notable is the observation that the perceptual intensity associated with the words was nearly identical for the two rating methods. The similarity of results that were obtained with the two rating methods supports the notion that subjects placed descriptors on the meterstick in a fashion that preserved the ratio scale qualities of those descriptors. In contrast, it may be that the essential structure of dissimilarity scaling is characterized by equal-interval categories. In this scenario, in spite of the instructions, the subjects may have evaluated the verbal descriptors in terms of equal-interval categories for both rating techniques. Finally, it is possible that the choice of verbal descriptors forced the appearance of an equal-interval scale. It should be noted again that the verbal descriptors were chosen to minimize rank order overlap. Since the degree of overlap between verbal descriptors was approximately constant across the various descriptors, the selection criteria would then tend to result in verbal descriptors that were equally spaced. As was discussed in the introduction, although many people have used magnitude estimation to judge odorant perceptual dissimilarity, the structure of the perceptual dissimilarity continuum is still elusive.

In the present experiment, the values attached to each of the adverbs were not grossly influenced by age or gender, which suggests that the resultant dissimilarity scale would be appropriate for most adults in judging levels of difference between stimuli. This was important because one of the uses of this scale will be to evaluate shifts in odorant quality in the hyposmic patient populations presenting to a State University of New York Health Science Center at Syracuse Smell and Taste Disorders Clinic. The segregation of young and old at age 50 was chosen because olfactory loss accelerates at about this age (Doty,

Figure 2. Perceptual intensities of descriptors obtained by magnitude estimation versus placement on a meterstick.

Shaman, & Dann, 1984). It should be noted that while our intent was to create a dissimilarity scale for rating odorant quality dissimilarity, the instructions to the subjects did not mention olfaction. Thus, it is suggested that the LDS would be an appropriate scale with which to judge dissimilarity over a wide range of stimuli and concepts.

ME has long been one of the widely accepted metrics of perceptual intensity (J. C. Stevens, 1971). Green et al. (1993) and Borg (1982) compared stimulus/response relationships generated with (1) ME and (2) their semantically labeled scale and found that their semantic scales possessed the same ratio scale properties as did ME, with little change in the slope of this function. Note, however, that metrics of odorant dissimilarity cannot be evaluated in the same way as metrics of perceptual intensity. Unlike in Borg's early work on taste intensity (Borg, Diamant, Strom, & Zotterman, 1967), there is no peripheral nerve from which to record perceptual dissimilarity and no stimulus continuum against which to plot perceptual difference. The metrics examined here assume that the general techniques of magnitude scaling provide an appropriate measure of dissimilarity, a position supported by a considerable body of literature (Berglund, Berglund, Engen, & Ekman, 1973; Gregson & Mitchell, 1974; Indow & Uchizono, 1960; Moskowitz & Barbe, 1976; Schiffman, Musante, & Conger, 1978; Schiffman, Nash, & Dackis, 1978; Schiffman et al., 1977).

As mentioned previously, the choice of descriptors and the placement of these descriptors on the LDS were determined without reference to a specific stimulus modality. This suggests the possibility that the LDS could be considered a general metric of perceptual dissimilarity. Although it is beyond the scope of the present work to present data to support this hypothesis, the extension of the LDS to another sensory system seems to be warranted. With perceptual dissimilarity's long history as a mainstay of visual psychophysics in mind, we evaluated the ability of the LDS to generate trichromatic color space.

EXPERIMENT 2

Color Dissimilarity—Generation of Color Space

Method

Subjects. Seven men and 5 women between the ages of 20 and 46 ($M = 31.67$; $SD = 7.00$) were recruited from the SUNY Health Science Center community (2 subjects had been in previous experiments). The 12 subjects were paid \$5 each for their participation.

Stimuli. Ten of the color caps (Nos. 8, 17, 25, 33, 42, 51, 59, 68, 76, and 84) from the Farnsworth–Munsell 100-hue color set (Gretag Macbeth) were chosen in such a way that the difference between one cap and the following cap comprised either eight or nine steps.

Procedure. Each subject was presented with all possible pairwise combinations of color caps (except self-matching) and was asked to rate the level of color difference between the members of the pairs with the LDS. Each of these ratings was performed in duplicate for a total of 90 ratings (45 per set). Each rating of color difference was made on a separate paper copy of the LDS. After a rating was made, the subject turned the rating over and proceeded to the next trial. Although testing was self-paced, all of the subjects completed testing within 30 min.

Results

Duplicate ratings from each subject were averaged and placed into a 10×10 dissimilarity matrix. These values were averaged across subjects for each pair of hues, and the results are shown in Table 4.

Table 4
Dissimilarity Matrix Obtained for Colors With the LDS

Color Cap	Color Cap									
	8	17	25	33	42	51	59	68	76	84
8	0.0									
17	35.7	0.0								
25	67.9	24.0	0.0							
33	74.8	47.5	13.1	0.0						
42	75.8	64.1	28.9	20.7	0.0					
51	78.8	78.0	54.2	46.2	19.8	0.0				
59	80.7	77.1	64.7	56.7	35.2	14.9	0.0			
68	73.6	70.0	74.1	69.0	59.9	51.1	36.7	0.0		
76	54.2	75.7	76.3	74.8	75.0	72.4	64.5	26.0	0.0	
84	29.3	62.5	75.6	81.7	76.1	78.9	79.0	61.2	29.0	0.0

Figure 3. Two-dimensional multidimensional scaling space created from color dissimilarities obtained with the labeled dissimilarity scale.

The dimensionality of the resulting 10×10 composite dissimilarity matrix was reduced by multidimensional scaling (Systat, Kruskal, monotonic, two dimensions) to produce a graphical representation of color space. The two-dimensional solution accounted for nearly all of the variance with minimal stress (proportion of variance = .99196, stress = 0.032). It is shown graphically in Figure 3.

Discussion

The color chips in the Farnsworth–Munsell 100-hue color set were matched for luminosity and varied only in hue. The resulting color space should be planar, an expectation that was fulfilled in the present experiment. Although the LDS was developed for the purpose of examining perceptual relationships between odorants, the generation of the color wheel with the LDS extends the usefulness of this metric to issues of dissimilarity in other sensory systems.

To be valuable, a sensory metric such as the LDS must be stable and reliable, generating the same rating for the same stimulus by the same subject, time after time. Ideally, consistency in scale usage across subjects would also allow judgments to be compared between subjects. Traditionally, comparisons of perceptions between subjects are enigmatic, because perceptions are intrinsically a within-subjects phenomenon. Borg (1982) addressed this prob-

lem with the assumption that maximal sensations produce equivalent perceptual magnitude across subjects. Therefore, he developed a scale that was anchored on the low end by 0 and on the high end by *maximal* sensation. Borg's theory of equivalency of maximal sensation (Borg, 1982) was assimilated into the scale developed by Green et al. (1993) as *strongest imaginable* and was also applied to the generation of the present measurement of dissimilarity (*most different imaginable*). On the basis of these theoretical issues, we expected that the LDS would exhibit a high level of stability both within subjects and across subjects. This was tested formally in Experiment 3.

An additional principle that formed the basis of Borg's (1982) and Green et al.'s (1993) work was that semantic labels were assumed to have similar meanings across subjects. That is, the meaning of *moderate* in a particular sensory system is stable across subjects and, in particular, more stable than an arbitrary number (as is the case in ME). The placement of semantic labels in appropriate locations on a scale assists the subjects in the representation of perceptual magnitude. The phenomenon of context refers to the susceptibility of ratings of a particular (set of) stimulus(i) to be altered by the context in which the ratings are made. Context effects can arise in the perceptual domain in such areas as adaptation and cross adaptation (Baylin & Moulton, 1975) or in the response domain

Table 5
Olfactory Stimuli in Experiment 3

Odorant	Solute	Solute Concentration
Set 1: Low Contrast		
Licorice	Trans-anethole	0.0019% v/v
Mint	R-carvone	6.3% v/v
Mothballs	Naphthalene	0.0063% w/v
Rose	Phenethyl alcohol	12.53% v/v
Rubbing OH	2-Propanol	12.53% v/v
Set 2: High Contrast		
Licorice	Trans-anethole	0.019% v/v
Mint	R-carvone	6.3% v/v
Mothballs	Naphthalene	0.0063% w/v
Rose	Phenethyl alcohol	12.53% v/v
Vinegar	Acetic acid	25% v/v

Note—All odorants were diluted with 1,2-Propanediol to a total volume of 15 ml.

(Zwislocki & Goodman, 1980). A metric that stabilizes shifts in the response domain, such as the stabilization provided by the semantic labels on the LDS, would be expected to reduce the effect of contrast convergence. In Experiment 3, we explored the prediction that the LDS should be less influenced by context effects than should ME.

EXPERIMENT 3

Comparison of the Stability of the LDS to ME

As mentioned previously, several studies have demonstrated that evaluations with labeled magnitude scales can produce ratings that have many of the same properties as ME and are also amenable to statistical manipulation (e.g., Green et al., 1996). The aim of Experiment 3 was to compare the LDS created in Experiment 1 and ME in terms of the stability of dissimilarity ratings (within and across subjects) and in terms of susceptibility to the effects of context.

Method

Subjects. After signing an IRB approved consent form, 41 subjects were screened with an UPSIT (Doty et al., 1984) for normal olfactory ability. One subject was excluded because of hyposmia (UPSIT scores less than 34). Forty subjects (14 men and 26 women) ranging in between 18 and 50 ($\bar{x} = 26.9$, $\sigma^2 = 8.1$) participated in the whole experiment (3 subjects had participated in previous experiments). All subjects were paid for their participation.

Stimuli. Each subject evaluated two sets of odorants (Table 5). Each odorant was presented to the subject in 30-ml amber glass bottles (Thomas Scientific, 9711E78). Note that for the purpose of this experiment, the concentration of acetic acid in the vinegar was elevated beyond that of commercially available vinegar, in order to produce an odorant with higher than normal pungency. On the basis of past experience, the vinegar was assumed to be considerably more dissimilar to the other odorants in the set than was rubbing alcohol.

Procedure. Each subject was presented with all possible pairwise combinations of odorants (except self-matching) and asked to rate the level of olfactory difference between the members of each pair. Each of these ratings was performed three times to increase reliability, for a total of 60 ratings (30 per set). The subjects were given a brief break (roughly 5 min) after rating 30 odorant pairs (one set). The order of presentation of the two odorant sets was counterbal-

anced across subjects, with all subjects receiving both sets. The presentation rate of the odorant pairs was self-paced; most of the subjects completed the experiment in less than 1 h.

The subjects were randomly assigned to one of two groups (20 subjects in each group) for testing. One group rated dissimilarity of the pairs of odorants with ME, whereas the other group rated odorant dissimilarity with an LDS. The subjects in both groups wrote their responses in booklets, one response per page. None of the subjects had access to ratings made previously in the experiment.

Magnitude estimation. The subjects were trained in ME through the evaluation of projected line lengths. In this training, the subjects were presented with lines of varying length and were asked to assign a number that described the subjective magnitude of the line length. For example, if one line was twice as long as another, a number should have been twice as large. After this training with line lengths, the subjects were asked to judge the degree of dissimilarity of the odor pairs in a similar fashion.

LDS testing. The subjects were asked to rate the degree of dissimilarity between the odorant pairs described above. As is typical of labeled magnitude testing, the subjects were not given practice, but were given the following instructions:

You will be presented with a series of odorant pairs. Your task is to determine how dissimilar the odors seem by marking along the scale partitioned by common verbal descriptors. First determine which descriptor most appropriately describes the difference between two odors, then fine-tune your rating by marking the scale at the proper location between that descriptor and the next most appropriate one. Thus, if you consider the difference best described by Moderately Different, but toward Mildly Different, place a mark below Moderately Different at the appropriate distance from Mildly Different. Rate the stimulus relative to other smell sensations you have experienced in daily life.

The subjects rated odorant dissimilarity on the LDS developed in Experiment 1 (see Figure 1). The last instruction was a reminder to the subjects that dissimilarities were to be rated in relation to their personal experiences and not in relation to imagined dissimilarities outside of these limits. This provided an upper bound for perceptual dissimilarities similar to that provided by Borg (1982) in his use of *maximal* [effort].

Results

The coefficients of variation (standard deviation divided by the mean) both within and across subjects were compared between the LDS and ME. To evaluate the ability of each metric to produce equivalent ratings under similar conditions, the coefficients of variation were calculated for each subject's replicate evaluations of each odorant pair. The coefficients within subjects were averaged across subjects for each stimulus, contrast order, and odorant set condition. The LDS was associated with a significantly (two-tailed paired t test, $p = .049$) but only slightly smaller variability ($CV = 0.38$) across replicate evaluations than was ME ($CV = .42$). This was a reflection of the fact that the LDS was associated with a higher frequency of small coefficients of variation and a lower frequency of high coefficients of variation than was ME.

Responses for ME were normalized in such a way that the average response across subjects for two scales was equated (thereby preserving the dispersion between subjects). The variability across subjects was tested formally by comparing the variation of the means for the two scales. The LDS resulted in variation across subjects (variance = 156) which was much smaller [$F(19,19) = 24.52$, $p < .001$] than that of ME (variance = 3,840.12).

Table 6
Pairwise Dissimilarity Between the Four Fixed Odorants and the Two Contrast Agents (Rubbing Alcohol and Vinegar) When Evaluated With the LDS and With ME

Odorant	LDS		ME	
	Rub Alc	Vinegar	Rub Alc	Vinegar
Licorice	55.5	71.5	14.5	23.6
Mint	56.8	77.1	11.9	23.6
Mothballs	52.0	70.2	12.7	20.2
Rose	54.5	77.1	12.5	24.4

The adjusted range for the average ME responses was between 7 and 200, whereas the range for average LDS responses was 30 to 74.

As indicated earlier, vinegar was substituted for rubbing alcohol to produce a high-contrast odorant set. The high contrast of vinegar was confirmed with both the LDS and ME. Ratings of dissimilarity between the four fixed members of the odorant set (rose, mothballs, licorice, and mint) were larger [one-tailed t test, $t(18) = 5.50$, $p < .001$ for the LDS, and one-tailed t test, $t(18) = 2.75$, $p < .01$ for ME] when these odorants were compared with vinegar than when compared with rubbing alcohol (Table 6, columns 2 vs. 3 for the LDS, columns 4 vs. 5 for ME). The ratings of dissimilarity within the four fixed members of the odorant set were 14% lower [one-tailed t test, $t(18) = 2.70$, $p < .01$ for ME] when the set contained the high-contrast vinegar odorant than when the set contained rubbing alcohol, the low-contrast odorant (Table 7, columns 2 and 3). Rating odorant dissimilarity with ME produced essentially the same results. The dissimilarity within the four fixed odorants was lower (Table 7, columns 4 and 5) when the odorant set contained vinegar than when the set contained rubbing alcohol [one-tailed t test, $t(18) = 1.74$, $p = .05$]. When evaluated with ME, the presence of vinegar in the odorant set resulted in a 25% reduction in perceived dissimilarity.

Discussion

In the present experiment, high and low contextual conditions were created by the substitution of vinegar for rubbing alcohol in the odorant sets and were confirmed in the ratings of the two metrics. Under the conditions of this experiment, both scales were sensitive to the effects of contrast convergence—a finding that is in agreement with Lawless's work in taste (Lawless, Horne, & Speirs, in press). However, contrary to Lawless's work, in the present experiment ME was found to be slightly more sensitive to contextual shifts than a labeled magnitude scale, in this case the LDS. As discussed previously, context effects can arise in either the perceptual domain or the response domain. The present finding that contrast effects are slightly smaller for the LDS is consistent with the possibility that this scale is less sensitive to shifts in the perceptual domain than is ME, a possibility that would suggest insensitivity rather than stability. However, these data are also consistent with the possibility that the LDS stabilized the response domain. It seems more likely that

the adjectives made it more comfortable for subjects to use the LDS and may have provided a slightly more stable conceptual framework in which to judge odorant dissimilarity.

The LDS produced dissimilarity ratings that were considerably more consistent across subjects than did ME. This consistency can be explained by the variety of scale ranges that were employed by people in the ME group as compared with the common-scale range provided by the LDS. Magnitude estimates are often normalized to reduce this variability. However, whereas normalization retains relative sensory magnitude, it eliminates possible interpretation of absolute sensory magnitude. It should be noted that the decision to remove *most different imaginable* from the scale was a bit risky. If, as Borg (1982) argued, the stability of his scale of perceived physical effort was mainly a result of equivalent perceptual intensities across subjects who were working at maximal physical effort, the stability of a labeled magnitude scale or LDS would be compromised by removing the top anchor of the scale. The stability of the LDS in the absence of *most different imaginable* suggests that adjectives and adverbs imply absolute levels of perceptual intensity. One of the benefits of labeled magnitude scales is that they are anchored by verbal descriptors that denote absolute levels of sensory magnitude (Green et al., 1993) and do not simply act as the perceptual multipliers explored in Cliff's (1959) work.

In contrast to intersubject variability, the LDS and ME showed similar levels of intrasubject variability. This is a bit surprising, since it might be assumed that the verbal labels on the LDS would help the subjects choose an appropriate scale location when rating stimulus pairs. However, it has been argued that subjects, in general, do not use numbers arbitrarily and that numbers themselves have perceptual magnitudes (Zwislocki & Goodman, 1980). As such, absolute ME carries with it references within each subject that provide a structure resulting in intrasubject consistency.

GENERAL DISCUSSION

As discussed in the introduction, the concept of perceptual dissimilarity has been useful in the description of normal and dysfunctional sensory systems. In olfac-

Table 7
Pairwise Dissimilarity Within the Four Fixed Odorants When Intermixed With the Two Contrast Agents (Low = Rubbing Alcohol, High = Vinegar) Evaluated With the LDS and With ME

Odorant	LDS		ME	
	Low	High	Low	High
Licorice-Mint	32.2	29.3	8.1	5.4
Licorice-Mothballs	55.8	47.4	13.0	9.9
Licorice-Rose	43.9	32.0	11.3	9.6
Mint-Mothballs	60.9	54.0	12.4	10.9
Mint-Rose	39.8	35.5	11.3	7.6
Mothballs-Rose	50.7	45.4	14.5	10.0

tion, most sensory testing has been directed toward quantification of the level of loss and/or to follow the time course of recovery. The most commonly utilized clinical tools have been odorant identification and threshold (Cain, Gent, Goodspeed, & Leonard, 1988; Costanzo & Zasler, 1991; Doty et al., 1984; Wright, 1987). However, a global measure of olfactory loss does not address odorant-specific hyposmias and parosmias, from which inferences could be made about the mechanisms of loss, or possibly olfactory coding (e.g., Kurtz, Hornung, Emko, & White, 1996; Kurtz et al., 1995; Kurtz et al., 1994; O'Connell et al., 1989; O'Connell et al., 1994; D. A. Stevens & O'Connell, 1991, 1996). For many years, the loss of specific odorant qualities has implied the existence of odorant-specific receptor molecules (Amoore, 1977). More recently, this view has been refined to suggest that odorant molecules interact with a number of perceptual channels or receptor molecules (D. A. Stevens & O'Connell, 1991, 1996). In each case, shifts in the odorant quality perception add to an understanding of the functioning of the olfactory system beyond that which would have been possible by simple measures of olfactory loss.

The LDS is a tool for the direct evaluation of perceived dissimilarity. It avoids the requirement for confusion matrices that stimuli be named. In olfaction, odors may be chosen to represent specific classes of chemicals, specific dissimilarities in odorant object space, or specific classes of olfactory dysfunction (e.g., odors of varying mucosal solubility or trigeminal vs. olfactory stimulants). Because the verbal descriptors were attached to the scale in a manner consistent with the estimated perceptual intensities generated within the context of perceptual ratios, the LDS might be assumed to produce ratio scale level data, a feature that is assumed to be important to the production of (odorant) object space with multidimensional scaling (Indow & Uchizono, 1960). Some work (Judd, 1967) suggests that ratio scale metrics overestimate perceptual dissimilarity, which results in a distortion of perceptual space. Although the shape of color space was preserved in Experiment 2, the shape of odorant object space is currently unknown. One test of whether a metric overestimates large perceptual differences is the *triangle inequality* rule, which states that the perceptual dissimilarity between stimuli A and B and between stimuli B and C must be at least as great as the dissimilarity between stimuli A and C (Gregson, 1972). In the present experiment, the dissimilarities rated with both the LDS and ME satisfy this rule. However, this rule may not apply to stimuli with separable attributes (Tversky & Gati, 1982), a feature that is not clearly the case for odorants.

The LDS provides a simple tool for collecting ratings of odorant dissimilarity that yields results that are comparable to those obtained with ME. In addition, it allows sensory magnitudes to be compared across subjects without the need for normalization to another stimulus domain (Marks, Borg, & Westerlund, 1992). It is readily accepted by subjects, who require little training prior to using the

scale. Furthermore, the stability and ease of presentation make the LDS an ideal metric of odorant dissimilarity in naive subjects.

REFERENCES

- AMOORE, J. E. (1977). Specific anosmia and the concept of primary odors. *Chemical Senses & Flavor*, **2**, 267-281.
- BAYLIN, F., & MOULTON, D. G. (1975). Adaptation and cross-adaptation to odor stimulation of olfactory receptors in the tiger salamander. *Federation Proceedings*, **34**, 388.
- BERGLUND, B., BERGLUND, U., ENGEN, T., & EKMEN, G. (1973). Multidimensional analysis of twenty-one odors. *Scandinavian Journal of Psychology*, **14**, 131-137.
- BIENFANG, R. (1941). Dimensional characterization of odors. *Chronica Botanica*, **6**, 249-250.
- 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). Amsterdam: North-Holland.
- BORG, G., DIAMANT, H., STROM, L., & ZOTTERMAN, Y. (1967). The relation between neural and perceptual intensity: A comparative study on the neural and psychophysical response to taste stimuli. *Journal of Physiology*, **192**, 13-20.
- BUCK, L. (1993). Identification and analysis of a multigene family encoding odorant receptors: Implications for mechanisms underlying olfactory information processing. *Chemical Senses*, **18**, 203-208.
- BUCK, L., & AXEL, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, **65**, 175-187.
- CAIN, W. S., GENT, J. F., GOODSPEED, R. B., & LEONARD, G. (1988). Evaluation of olfactory dysfunction in the Connecticut Chemosensory Clinical Research Center. *Laryngoscope*, **98**, 83-88.
- CLIFF, N. (1959). Adverbs as multipliers. *Psychological Review*, **66**, 27-44.
- COSTANZO, R. M., & ZASLER, N. D. (1991). Head trauma. In T. V. Getchell, R. L. Doty, L. M. Bartoshuk, & J. B. Snow, Jr. (Eds.), *Smell and taste in health and disease* (pp. 711-730). New York: Raven Press.
- DAVIS, R. G. (1979). Olfactory perceptual space models compared by quantitative methods. *Chemical Senses & Flavor*, **4**, 21-33.
- DOTY, R. L., SHAMAN, P., & DANN, M. (1984). Development of the University of Pennsylvania Smell Identification Test: A standardized microencapsulated test of olfactory function. *Physiology & Behavior*, **32**, 489-502.
- FARNSWORTH, D. (1943). The Farnsworth-Munsell 100-hue and dichotomous tests for color vision. *Journal of the Optical Society of America*, **33**, 568-578.
- GESCHIEDER, G. A. (Ed.) (1976). *Psychophysics: Method and theory*. New York: Wiley.
- GREEN, B. G., DALTON, P., COWART, B., SHAFFER, G., RANKIN, K., & HIGGINS, J. (1996). Evaluating the labeled magnitude scale for measuring sensations of taste and smell. *Chemical Senses*, **21**, 323-334.
- GREEN, B. G., SHAFFER, G. S., & GILMORE, M. M. (1993). Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chemical Senses*, **18**, 683-702.
- GREGSON, R. A. M. (1972). Odour similarities and their multidimensional metric representation. *Multivariate Behavioral Research*, **7**, 165-174.
- GREGSON, R. A. M., & MITCHELL, M. J. (1974). Odor quality similarity scaling and odor-word profile matching. *Chemical Senses & Flavor*, **1**, 95-101.
- INDOW, T., & UCHIZONO, T. (1960). Multidimensional mapping of Munsell colors varying in hue and chroma. *Journal of Experimental Psychology*, **59**, 321-329.
- JUDD, D. B. (1967). Interval scales, ratio scales, and additive scales for the sizes of differences between members of a geodesic series of colors. *Journal of the Optical Society of America*, **57**, 380-386.
- KEMP, S. E., ISHII, R., & O'MAHONY, M. (1996, August). *Representation of odor quality on a spherical surface in three-dimensional stim-*

- ulus space. Presented at European Chemoreception Research Organization XII, Zurich.
- KURTZ, D. B., HORNUNG, D. E., EMKO, P. E., & WHITE, T. L. (1996). Relative odorant identification in the evaluation of hyposmia. *Chemical Senses*, **21**, 628.
- KURTZ, D. [B.], HORNUNG, D. [E.], KENT, P. [F.], EMKO, P. [E.], WHITE, T. [L.], & SHEEHE, P. [R.] (1995). Odorant dissimilarity as a measure of changes in the human olfactory system: Implications for the evaluation of dysosmia. *Chemical Senses*, **20**, 726-727.
- KURTZ, D. B., SHEEHE, P. R., KENT, P. F., HORNUNG, D. E., WHITE, T. L., YOUNGENTOB, S. L., SCHWOB, J. E., MOZELL, M. M., BELKNAP, E. M., EMKO, P. E. (1994). Odorant confusion matrix response patterns: Relations to medical history. *Chemical Senses*, **19**, 501.
- KUZNICKI, J. T. (1980). Magnitude estimates of distances along category rating scales. *Perceptual & Motor Skills*, **50**, 147-150.
- LAWLESS, H. T., HORNE, J., & SPEIRS, F. (in press). Contrast and range effects for category, magnitude and labeled magnitude scales. *Chemical Senses*.
- MARKS, L. E., BORG, G., & LJUNGGREN, G. (1983). Individual differences in perceived exertion assessed by two new methods. *Perception & Psychophysics*, **34**, 280-288.
- MARKS, L. E., BORG, G., & WESTERLUND, J. (1992). Differences in taste perception assessed by magnitude matching and by category-ratio scaling. *Chemical Senses*, **17**, 493-506.
- MOSKOWITZ, H. R., & BARBE, C. D. (1976). Psychometric analysis of food aromas by profiling and multidimensional scaling. *Journal of Food Science*, **41**, 567-571.
- O'CONNELL, R. J., STEVENS, D. A., AKERS, R. P., COPPOLA, D. M., & GRANT, A. J. (1989). Individual differences in the quantitative and qualitative responses of human subjects to various odors. *Chemical Senses*, **14**, 293-303.
- O'CONNELL, R. J., STEVENS, D. A., & ZOGBY, L. M. (1994). Individual differences in the perceived intensity and quality of specific odors following self- and cross-adaptation. *Chemical Senses*, **19**, 197-208.
- SCHIFFMAN, S. S. (1974). Contributions to the physiochemical dimensions of odor: A psychophysical approach. In W. S. Cain (Ed.), *Odors, evaluation, utilization, and control* (Annals of the New York Academy of Sciences, Vol. 237, pp. 164-183). New York: New York Academy of Sciences.
- SCHIFFMAN, S. S., MUSANTE, G., & CONGER, J. (1978). Application of multidimensional scaling to ratings of foods for obese and normal weight individuals. *Physiology & Behavior*, **21**, 417-422.
- SCHIFFMAN, S. S., NASH, M. L., & DACKIS, C. (1978). Reduced olfactory discrimination in patients on chronic hemodialysis. *Physiology & Behavior*, **21**, 239-242.
- SCHIFFMAN, S. S., ROBINSON, D. E., & ERICKSON, R. P. (1977). Multi-dimensional scaling of odorants: Examination of psychological and psychochemical dimensions. *Chemical Senses & Flavor*, **2**, 375-390.
- SHEPARD, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function: II. *Psychometrika*, **27**, 219-246.
- STEVENS, D. A., & O'CONNELL, R. J. (1991). Individual differences in thresholds and quality reports of human subjects to various odors. *Chemical Senses*, **16**, 57-67.
- STEVENS, D. A., & O'CONNELL, R. J. (1996). Pemenone and androstene do not cross-adapt reciprocally. *Chemical Senses*, **21**, 711-717.
- STEVENS, J. C. (1971). [Book review of *Sources of Color Science*, ed. by David L. MacAdam]. *Contemporary Psychology*, **16**, 564-565.
- TVERSKY, A., & GATI, I. (1982). Similarity separability and the triangle inequality. *Psychological Review*, **89**, 123-154.
- WOSKOW, M. H. (1966). Multidimensional scaling of odors. In NATO (Ed.), *NATO Advanced Study Institute on Theories of Odor and Odor Measurement* (pp. 147-188). Istanbul: NATO.
- WRIGHT, H. N. (1987). Characterization of olfactory dysfunction. *Archives of Otolaryngology*, **113**, 163-168.
- YOSHIDA, M. (1964). Studies of psychometric classification of odors (5). *Japanese Psychological Research*, **6**, 145-154.
- ZHENG, C., HOFFMAN, H. J., LUCCHINA, L. A., BARTOSHUK, L. M., & WEIFFENBACH, J. M. (1998). Comparison of the Green scale versus magnitude estimation for taste perception. *Annals of the New York Academy of Science*, **855**, 820-822.
- ZWISLOCKI, J. J., & GOODMAN, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, **28**, 28-38.

(Manuscript received November 20, 1997;
revision accepted for publication October 3, 1998.)