Contrast and Range Effects for Category, Magnitude and Labeled Magnitude Scales in Judgements of Sweetness Intensity

Harry T. Lawless, John Horne and Ward Spiers

Cornell University, Ithaca, New York, NY 14853, USA

Correspondence to be sent to: Harry Lawless, Department of Food Science, Stocking Hall, Cornell University, Ithaca, New York, NY 14853, USA. e-mail: htl1@cornell.edu

Abstract

The labeled magnitude scale (LMS) is a verbally anchored quasi-logarithmically spaced response scale with properties similar to magnitude estimation. Three experiments examined whether the LMS showed context effects similar to those found with magnitude estimation and category scales. Two versions of the LMS were used, one anchored at the high end to the strongest imaginable sweetness and the other to the strongest imaginable oral sensation. In a simple contrast experiment, subjects judged the sweetness of a 10% sucrose fruit beverage in the context of a less sweet (5%) beverage or a more sweet (20%) beverage. Consistent with previous literature, the sweetness was judged more intense in the low context and less intense in the high context, for all scaling methods. In a second experiment, this effect persisted (although was smaller) when the contextual item preceded the to-be-rated item, a so-called 'reversed-pair' design. Once again, the effect was highly significant for all scaling methods. In a third experiment, a range effect was examined using wide and narrow ranges of concentration. Psycho-physical functions were flatter in a wide context and steeper in a narrow context, consistent with previous observations on range-mapping bias. This result was obtained for all scales. In three common contextual effects, the labeled magnitude scale behaved similarly to other scaling procedures. Its application to comparisons across individuals may be limited if those individuals have different experiential contexts within which they make their judgements.

Introduction

The labeled magnitude scale is a method used for obtaining intensity estimates. Developed by Green and colleagues (Green *et al.*, 1993, 1996), the method consists of a line (usually vertical) with verbal labels (weak, moderate, strong, etc.) for intensity levels spaced in a quasi-logarithmic fashion (Figure 1). Subjects generally are instructed to place a mark on the line to reflect the perceived intensity of a stimulus. The method produces psychophysical functions similar to those produced by magnitude estimation (Stevens, 1956).

Based on earlier work by Borg and colleagues (Marks *et al.*, 1983; Borg, 1992), the upper-end anchor of this scale is labeled 'strongest imaginable'. The question arises as to whether this upper-end anchor provides an extra-experimental and stable frame of reference for subjects, or whether the experimental context is capable of overriding any stabilization provided by anchoring to an intense or imaginably intense life experience. A related question is whether such experiences provide a frame of reference that is common between individuals, thus facilitating comparisons of intensity ratings among individuals or between groups, such as those that differ in bitterness sensitivity to PROP (propyl-thiouracil). In studying the psychophysics of perceived exertion, Borg proposed that exercising to the maximum of one's exertion would provide a common subjective experience. If such a common frame of reference could be achieved, the scale would seem to have wide applications in studies of individual differences in sensory function, as well as in clinical studies of various patient populations.

The following experiments were conducted to test whether the labeled magnitude scale was more or less prone to the common contextual biases produced by contrasting stimuli and stimulus range. The immediate stimulus context in a scaling study exerts a strong influence on intensity ratings assigned to a particular stimulus (Helson, 1964, Poulton, 1989; Schifferstein and Frijters, 1992). Stimuli are often contrasted with those immediately preceding, a natural human tendency to compare items as opposed to reflecting on their intensity in some absolute sense. This tendency renders most (if not all) scaling methods relative in nature, as if subjects were using a kind of rubber ruler that would stretch or contract to fit the immediate situation (Riskey *et al.*, 1979). To the extent that psychophysical intensity scaling is relative to context, any differences in the frames of reference that different individuals use would make comparisons between such individuals risky at best.
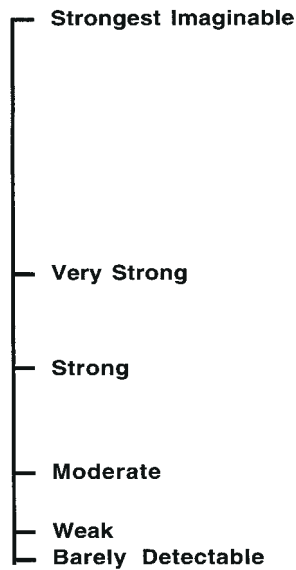
Figure 1 The labeled magnitude scale after Green *et al.* (Green *et al.*, 1993).

Experiment 1—a simple contrast effect

A common finding in intensity judgements is one of contrast. Contrast occurs when a target stimulus is shifted away in comparison to a contextual stimulus. For example, a stimulus is judged to be weaker in the presence of stronger stimuli and stronger in the presence of weaker stimuli (Lawless, 1983; Mattes and Lawless, 1985; Poulton, 1989; Rankin and Marks, 1991). The reverse effect of assimilation, shifting in the direction of the contextual stimulus, is also sometimes seen (Ward, 1982; Wedell, 1990). Contrast effects have been observed with a variety of response tasks, including category ratings (Helson, 1964), magnitude estimation (Mellers and Birnbaum, 1982), open-ended written descriptions (Simpson and Ostrom, 1976), line marking scales (Lawless and Malone, 1986) and cross-modality matching procedures (Manis, 1967). Contrast may arise from a pervasive human tendency to compare stimuli to the current frame of reference. In Ithaca, NY, a January day that reaches 40°F is considered quite mild and pleasant, while an evening in early September that dips to 40°F feels quite cool. Contrast effects are also seen in hedonic ratings (Schifferstein, 1995) and in perceptual quality judgements. For example, terpene odors with partially citrus and partially woody character are judged to be more citrus-like in the presence of woody type odors and are judged to be more woody in the presence of citrus-like odors (Lawless *et al.*, 1991). This experiment tested whether or not the contrast effects seen with category scales and magnitude estimation were similar to those that might be observed with the LMS method.

Methods

Twenty-five subjects (12 female) were recruited from the Cornell community and were paid for participation in the study.

Stimuli consisted of grape flavored Kool-Aid sweetened with sucrose to three different levels: 5, 10 and 20% (w/v). The sucrose was stirred into solution in room temperature spring water to bring the total volume to 2 l. The flavorant from one envelope of Kool-Aid (3.9 g) was then added to the mixture. Test solutions were poured into 20 ml samples in plastic cups labeled only with random three digit codes and were stored at room temperature. Solutions were first used no earlier than 1 day after being prepared and were replaced, if necessary, 4 days after being prepared.

Judgements were of perceived sweetness intensity only. Four scaling methods were used as follows: (i) a 15-point category scale anchored at either end with the phrases 'not sweet' (left-most box) and 'extremely sweet' (right-most box); (ii) a magnitude estimation scale in which the reference sample (10% sucrose w/v) was assigned a rating of 10; (iii) a labeled magnitude scale in which subjects were instructed to rate the sweetness of the stimuli relative to the strongest sensation of sweetness they could imagine; and (iv) a labeled magnitude scale in which subjects were instructed to rate the stimuli relative to the strongest oral sensation they could imagine, discounting sensations associated with pain. Exact wording of the instructions on each of the ballots was similar to that used by Green *et al.* (Green *et al.*, 1996). The labeled magnitude scales consisted of a bold vertical line 135 mm in length (the same length as the horizontal category scale), with six verbal anchors at the following scale points: 'barely detectable' (2 mm), 'weak' (9 mm), 'moderate' (23 mm), 'strong' (48 mm), 'very strong' (72 mm) and 'strongest imaginable' (135 mm). Subjects were instructed to make their rating by placing a mark anywhere on the line. Ratings were scored in millimeters from the bottom of the scale.

For each of the ballots the subjects tasted and rated the same sets of stimuli in both high and low context conditions. A context stimulus was followed by a test stimulus. The context stimulus was either the 20% sucrose mixture (high context) or the 5% sucrose mixture (low context); the test stimulus was always 10% sucrose. For the magnitude estimation procedure, subjects tasted the 10% sucrose mixture as a reference stimulus prior to tasting the context stimulus. Room temperature spring water was available for rinsing. They were verbally instructed to rinse between replicates and ballots. Subjects participated in four experimental sessions. In each session, subjects were presented with the stimuli for two of the scaling techniques in the same context condition. The order of presentation of the scaling techniques and the order of context condition presentation was randomized across subjects. Three replicates were obtained for each scaling technique and context condition. Replicates were obtained within the same experimental sessions.

Data were analysed using SYSTAT v. 5.2. Means from the individual scaling techniques were treated independently

from each other in repeated measures ANOVAs with context condition, replicates and subjects (a random effect) as factors. The data from each scaling technique were then transformed to percentages of scale ranges. A rating of 40 was used as the upper boundary for the magnitude estimation scale, since that rating was the highest rating less than three standard deviations above the mean for both context conditions. A few outlier data points were therefore recoded to 100% of scale range for this scale only. The transformed data were then subjected to a single repeated measures ANOVA with scales, context, conditions, replicates and subjects (random) as factors.

Results and discussion

Figure 2 shows the mean ratings for the four scaling methods, two contexts and replicates. For all methods, the context effect was one of contrast. The 10% sucrose target stimulus was rated higher when the 5% sucrose was presented and lower when the 20% sucrose was presented [main effect of context, $F(1,24) = 87.06$, $P < 0.001$]. There was also a significant scale by context interaction [$F(3,72) = 13.38$, $P < 0.001$]. The contextual shift appears to be somewhat smaller for magnitude estimation and largest for the category scale. A replicate by context interaction [$F(2,48) = 8.84$, $P < 0.001$] is consistent with the pattern of increasing contrast across replicates as the context became learned and the perceptual frame of reference solidified (Lawless, 1983). In summary, all scales were prone to the contextual effects of contrast.

Experiment 2—a reversed-pair contrast effect

Experiment 2 was conducted to test the generality of the simple contrast effects seen in Experiment 1 using a reversed-pair experiment. In the reversed-pair experiment, the contextual stimulus follows the target stimulus, which is then rated from memory. Due to the order of context and target, a contrast effect that persists in the reversed-pair situation can not be attributed to simple sensory adaptation or other peripheral sensory encoding processes (Diehl *et al.*, 1978; Lawless *et al.*, 1991). If a context effect is eliminated by the reversed-pair presentation, it is likely that the contextual influence was an early-on adaptation-like process, one exerted before its dimensional placement or response translation, as discussed further below (Wedell, 1990).

Methods

Twenty-five subjects (12 female) were recruited from the Cornell campus. Stimuli were the same as those used in Experiment 1, except that the flavorant used was changed to raspberry, which required 3.6 g (1 envelope) per 2 l of water. Mixture, storage and serving conditions were also the same as those described above.

Scaling methods and instructions were the same as in
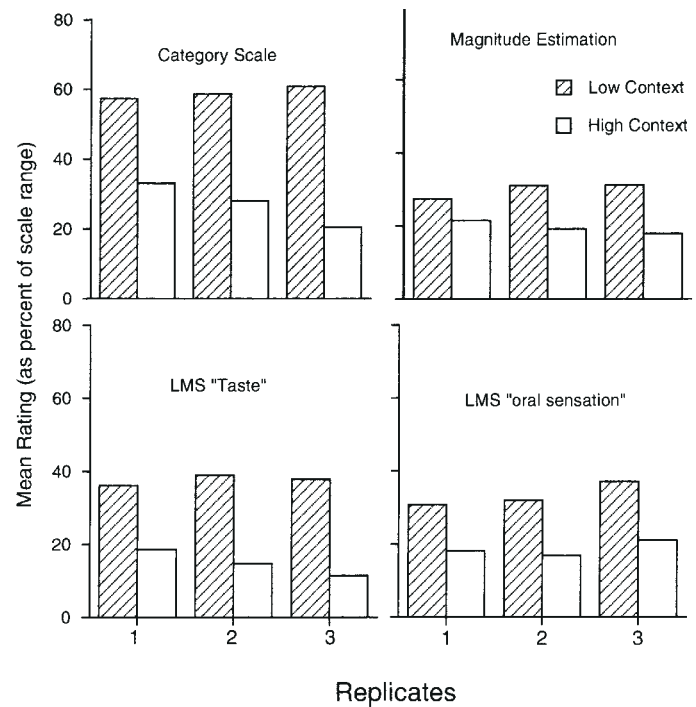
Figure 2 Mean ratings in the high and low context conditions for the four methods, plotted across replicates.

Experiment 1. Subjects were exposed to all four methods in two context conditions. Three replicates of each condition were obtained in the same experimental session. Experiment 2 differed from Experiment 1 only in the order of presentation of the stimuli. In Experiment 2, subjects were first exposed to the test stimulus (10% sucrose) followed by the context stimulus (5% sucrose in the low context condition and 20% sucrose in the high context condition). Subjects were then asked to rate the sweetness of the test stimulus by memory without retasting it. Subjects were told 'This is a study to determine how well people can remember what certain foods taste like' and then, following the second, contextual stimulus, 'Please rate the sweetness of the first sample you tasted in this series. Remember, you are rating the first sample. Do not go back and re-taste the sample before you rate it.' For the magnitude estimation procedure, subjects tasted the 10% sucrose mixture as a reference stimulus prior to tasting the test stimulus (which was then followed by the contextual stimulus). This was assigned a sweetness 'rating' of 10.

Data were analysed using SYSTAT v. 5.2. Means from the individual scaling techniques were treated independently from one another in repeated measures ANOVAs with context condition, replicates and subjects (a random effect) as factors. The magnitudes of the context effects found in Experiment 1 were compared with those in Experiment 2 by representing the context effect in individual studies as a single difference score (i.e. 10% sucrose rated under the low context condition minus 10% sucrose rated under the high

Table 1 Mean ratings, reversed-pair experiment

	Replicate 1	Replicate 2	Replicate 3
Category scale			
High context	6.28	5.48	5.32
Low context	8.16	8.96	8.60
Magnitude estimation			
High context	9.32	9.24	8.60
Low context	12.68	11.80	12.62
LMS (strongest taste)			
High context	25.04	21.21	21.88
Low context	37.16	36.44	36.32
LMS (strongest oral sensation)			
High context	24.60	22.84	21.80
Low context	31.60	37.72	37.04

Figure 3 Mean difference scores (low context condition minus high context condition) for the simple contrast and reversed-pair experiments, converted to percentage of scale range.

context condition). These difference scores were calculated for each individual using each scaling technique in both experiments. Means were calculated for the individual scaling techniques and compared via repeated measures ANOVA.

Results and discussion

Table 1 shows the mean values in the two contexts for the four scaling methods across replicates and a final across-replicate average. As in Experiment 1, there was a significant contrast effect, with stimuli judged sweeter in the context of the 5% sample and less sweet in the context of the 20% sample [$F(1,24) = 64.03$, $P < 0.001$]. Results were more stable across replicates in this design, with no replicate effects or interactions. There was a significant difference among scales once the data were converted to percentage of scale range, with the category scales generally higher than the other three methods [$F(3,72) = 2.80$, $P < 0.05$].

Data were then converted to difference scores to compare the effect size from the first and second experiments. Figure 3 shows the mean difference scores as a function of each scaling type and experimental design. The contextual shift was larger in the context-first design as opposed to the reversed-pair design [overall, $F(1,24) = 8.75$, $P < 0.01$]. The fact that the contextual shift persisted in the reversed-pair design is consistent with the placement of the contrast effect at the level of dimensional placement or response translation, rather than at an initial sensory encoding process (Wedell, 1990). However, the decrement in the size of the contrast shift could reflect a partial effect at the sensory encoding stage in Experiment 1, such as simple sensory adaptation.

Experiment 3—stimulus range effects

Range effects are commonly observed as a flattening of the psychophysical function when wider stimulus ranges are presented (Engen and Levy, 1958; Teghtsoonian and
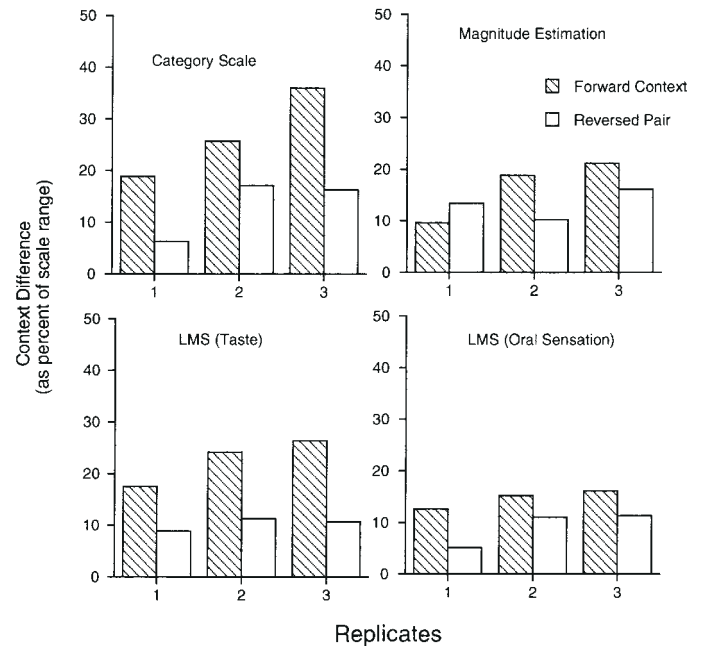
Teghtsoonian, 1978; Poulton, 1989). In ratio-scaling studies, this is often seen as a lowering of the power function exponent or a flattening of the slope in a logarithmic plot. However, range effects are common to many scaling methods (Parducci, 1965). For example, Lawless and Malone (Lawless and Malone, 1986) found steepened psychophysical functions for category, line marking and magnitude estimation methods when stimulus range was truncated. The effect was quite profound, with an increasing category scale usage of from three to six points on a nine-point scale (33% of scale range) for one stimulus range in texture judgements [see p.172, figure 8 of (Lawless and Malone, 1986)]. The third experiment was conducted to see if similar shifts in scale usage would occur for the LMS.

Methods

The subjects included 15 males and 13 female students at Cornell University. They were untrained except for instructions given in this study on scale usage. None had participated in the earlier studies and they were unaware of the purpose or predictions of this experiment. Subjects were paid for participation after completion of the four sessions.

Stimuli consisted of six samples of a grape-flavored powdered drink mix with added sucrose. In the wide stimulus range, samples consisted of 2.5, 8, 12 and 25% sucrose (w/v) and in the narrow range condition the samples were 5, 8, 12 and 17.5% sucrose. A 10% sucrose sample served as the reference in the magnitude estimation condition.

Four sessions were conducted on four consecutive days, two with narrow ranges and two with wide ranges. Two

scaling methods were administered per day with replicate judgements of each concentration. The orders of scaling methods and ranges were counterbalanced among subjects. Sessions were performed in a quiet secluded room under fluorescent light with panelists separated by at least 15 feet to insure independent judgements. Subjects were instructed not to consume beverages or food or to chew gum for 30 min before participating.

Judgements were of perceived sweetness intensity only. Scaling methods included: a 15-point category scale using check boxes with end anchors labeled from 'not sweet at all' on the left-most box to 'extremely sweet' on the right-most box; magnitude estimation with a 10% sucrose reference being assigned a value of 10; a labeled magnitude scale with the upper end anchored to 'the strongest oral sensation ever experienced' [called LMS(o)]; and a labeled magnitude scale with the upper end anchored to 'the strongest sweet sensation' ever experienced by the subject [called LMS(s)]. The labeled magnitude scales consisted of a 135 mm bold vertical line with the following labels and placements: 'barely detectable' (2 mm), 'weak' (9 mm), 'moderate' (23 mm), 'strong' (48 mm), 'very strong' (72 mm) and 'strongest imaginable' (135 mm). Ratings were scored as distance measured from the bottom of the scale and converted to a 1–100 range.

Data were analysed using SYSTAT 5.2 with paired *t*-tests and sign tests used to compare slope fit to individual functions as described below.

Results and discussion

Linear functions were fit for each subject to the category data as a function of log concentration, and to the log data of the other three scaling methods, also as a function of log concentration. This was done to conform to the Fechnerian (semilog) function for category scale data and the Stevens power function (linear in a log–log plot) for the magnitude estimation (ME) and LMS data. Data were averaged across replicates for each individual to smooth the functions before curve fitting (least squares method). The fitted functions took the form

$$\text{mean rating (category)} = k_1 \log(\text{concentration}) + k_2$$

or

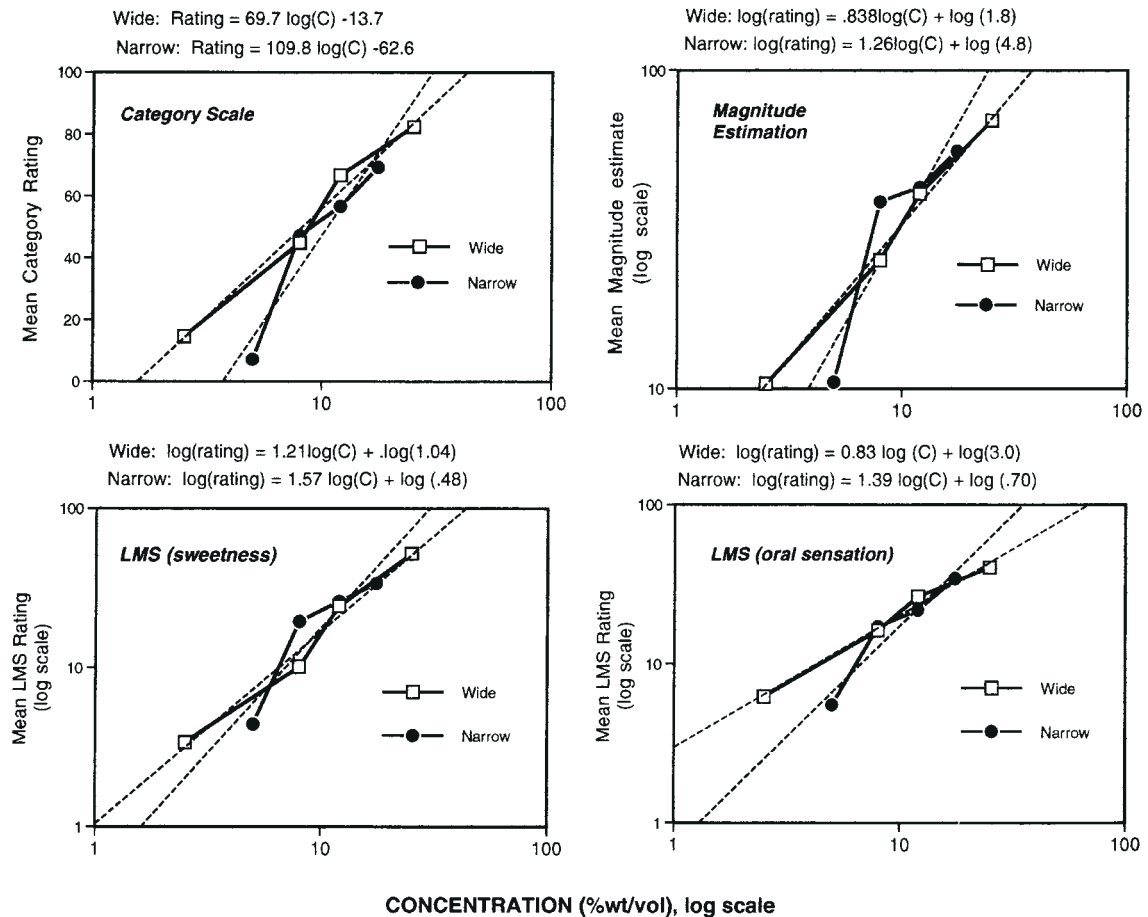$$\log(\text{mean rating (ME, LMS(s), LMS(o))}) = \\ k_1 \log(\text{concentration}) + k_2.$$

A simple test of the range effect then becomes a comparison of slopes, k_1, for these respective equations, since the range effect predicts a steeper slope in the narrower range (consistent with the expansion of rating scale usage to map the sensations onto the available scale range). Although there was some nonlinearity in the individual scaling data with only two replicates, the functions produced a reason-ably good fit to the data. Adjusted R^2 values were >0.5 (in both ranges) for 25, 23 and 28 of 28 subjects, for the category scales, magnitude estimation and LMS(s) scales, respectively. A higher level of error variation and poorer fits were found to the LMS scale data anchored to 'strongest oral sensation' as discussed below.

Results were similar for three of the four methods. The category scaling showed a mean slope of 9.83 (±0.41, standard error of the mean) for the wide range and 15.46 (±0.73) for the narrow range [$t(27) = 6.92$, $P < 0.001$], with 25 of 28 subjects changing slope in the predicted direction (sign test, $P < .0001$). The magnitude estimation scaling produced a mean slope (power function exponent) of 0.903 (±0.06) for the wide range and 1.421 (±0.10) for the narrow range [$t(27) = 6.12$, $P < 0.001$], with 25 of 28 subjects changing slope in the predicted direction (sign test, $P < 0.001$). For the LMS(s) scale, the slope in the wide range was 1.23 (±0.058) and the slope in the narrow range was 1.69 (±0.094) [$t(27) = 5.97$, $P < 0.001$], with, again, 25 of 28 subjects changing in the predicted direction (sign test, $P < 0.001$).

The change to steeper psychophysical functions in the narrow range was also observed for the LMS scale anchored to 'strongest oral sensation'. However, six subjects had difficulty using this scale when rating sweetness relative to oral sensation. Three produced negative or near-zero slopes (adjusted $R^2 < 0.1$) in both conditions and three others in one of the two range conditions. Inspection of the raw data showed significant reversals with concentration or non-discriminating flat functions. Their data were eliminated from the analysis below, although their inclusion would not have changed the conclusions, statistical outcome or probability levels. Based on the 22 remaining subjects, the slope in the wide range was 0.888 (±0.060) and 1.552 (±0.119) in the narrow range [$t(21) = 4.88$, $P < 0.01$], with 19 of 22 subjects showing a shift in the predicted direction (sign test, $P < 0.001$). The slightly lower slopes with anchoring to the strongest sensation are consistent with the observation of Green *et al.* (Green *et al.*, 1996) that the LMS produces slightly expanded ratings when anchored to a taste frame of reference.

Figure 4 shows the mean values for each rating method and range condition plotted as a function of log concentration. Category scales are shown in semilog and the other three methods in log–log plots. Ratings have been converted to percentage of scale range for comparison. Since magnitude estimation is open ended, a functional scale range maximum was determined from the overall mean score plus two standard deviations (essentially the 97.5 percentile, eliminating some high outliers common to this scaling method). All four methods show the common range effect of steeper slopes in narrower stimulus ranges. As in the analysis of individual curves, the increase in slope from the wide to narrow stimulus range was ~30–50%.

Wide: Rating = 69.7 log(C) -13.7
Narrow: Rating = 109.8 log(C) -62.6

Wide: log(rating) = .838log(C) + log (1.8)
Narrow: log(rating) = 1.26log(C) + log (4.8)

Wide: log(rating) = 1.21log(C) + .log(1.04)
Narrow: log(rating) = 1.57 log(C) + log (.48)

Wide: log(rating) = 0.83 log (C) + log(3.0)
Narrow: log(rating) = 1.39 log(C) + log (.70)

CONCENTRATION (%wt/vol), log scale

Figure 4   Mean ratings in wide and narrow stimulus ranges for the four methods. Dashed lines show the fitted functions (linear in semilog for the category scales and linear in log log for the other three methods).

General discussion

Muñoz and Civille (Muñoz and Civille, 1998) classified different approaches to scaling as they are used in applied descriptive evaluations of foods as universal, product specific or attribute specific. These three scale types are used in making a descriptive profile of all the sensory attributes in complex food products. The critical differences among these methods concern the nature of how the scales are anchored to intensity references. 'Product specific' scales are anchored to the most intense sensory attribute in the product set determining the upper bound. All attributes are rated relative to this intensity level. This renders comparisons between scale values for different products feasible, much like cross-modality matching or magnitude matching studies in psychophysics (Stevens and Marks, 1980). In this case it would be permissible to speak of 'the sweetness of a product as being twice as intense as the saltiness'. In 'attribute specific' scales, each attribute or characteristic has its own upper reference, effectively stretching each scale to make it fully usable across the entire range for a given class of products, but rendering inter-attribute comparisons im-

possible since each has its own different upper intensity reference. Thus a cookie might be rated as both extremely sweet and extremely salty for such cookies, but this would not necessarily imply that the sweetness and saltiness were equi-intense, nor would it imply that the cookie was as salty as a potato chip. The third approach to scaling in industrial applications is the use of 'universal scales'. Universal scales are anchored to physical references provided during training to calibrate panelists to have common scale usage in product evaluations. The intensity frame of reference is calibrated across all perceptual attributes, particularly with reference to how the high end of the scale is to be interpreted.

The labeled magnitude scale could be viewed as another example of a universal scale in principle, because the upper-end anchor is the strongest sensation imaginable. However, Green *et al.* (Green *et al.*, 1996) showed that anchoring the scale to the strongest imaginable taste would produce results slightly different from the strongest imaginable oral sensation. Given that oral sensations, such as the trigeminally mediated chili burn, can be absolutely more intense than most tastes, it is reasonable that scale usage would expand for the subjectively truncated range of tastes.

This is analogous to Poulton's range mapping bias, in which the scale usage expands to fit the available stimuli and/or the resulting sensation range (Poulton, 1989). This was one early source of evidence that the LMS might be subject to the sorts of contextual shifts in intensity judgements that were well-known for other scaling techniques.

The present results are consistent with the idea that the labeled magnitude scale behaves very much like other scaling methods with respect to its susceptibility to common contextual effects of range and contrast. Anchoring to a concept of intense oral or taste sensations does not appear to protect the subject in any way from showing the kinds of contextual shifts seen in other scaling methods. Psycho-physical researchers should be cautious in assuming that use of the LMS will facilitate inter-group comparisons (such as between groups with different bitter sensitivities or between capsaicin-desensitized and non-desensitized individuals), especially if those groups have different perceptual contexts or frames of reference.

The robust nature of these contextual effects with un-trained subjects suggests that the perceptual shifts seen here are a function of processes that may precede the final rating assignment process in a scaling task. Wedell (Wedell, 1990) discussed the locus of various context effects within the flow of information. He distinguished between peripheral sensory effects, such as adaptation and perceptual effects, that reflected the placement of an encoded stimulus along a perceptual continuum (such as a perceived intensity con-tinuum), termed 'dimensional placement'' and response-related or judgemental effects having to do with cognitive selection of an overt response alternative.

The ubiquitous nature of contrast effects in a variety of numerical and non-numerical response tasks would seem to argue against a simple response translation effect, at least as the sole cause of contextual shifting. The persistence of contrast shifts in the reversed-pair experiment is consistent with the notion that the effects are not at the level of initial sensory encoding, but rather at the level of dimensional placement or beyond (Wedell, 1990). However, the reduction in the size of the effect from the forward-context experiment to the reversed-pair experiment suggests both a sensory/perceptual and a judgemental component. This presumes that the reduction in the effect during reversed-pair presentation is not due merely to an increased memory load or increased time intervals. Although the present studies can not determine the exact locus of origin or mechanisms giving rise to context and range effects, several alternative explanations have been proposed in the literature.

A well-known judgement-based theory of contextual shifting is the range-frequency theory of Parducci (Parducci, 1965). The frequency principle predicts that overutilization of high categories in an intense stimulus set would lead to a shifting downward in order to use lower response alterna-tives. Conversely, overuse of low responses in a low-intensity context would cause a shift upward in ratings. This could give rise to the common contrast effect wherein an intermediate stimulus is viewed as less intense in a high series and more intense in a low series. However, both con-trast and assimilation have been observed under different conditions, and a judgemental or response driven effect is not the only explanation for contextual shifting. Ward (Ward, 1982), in studying stimulus and response depend-encies in sequences, found contrast among stimuli but assimilation among responses. This is consistent with assimilation in response-related processes and contrast occurring earlier in stimulus encoding. Along these lines, Marks (Marks, 1992) argued that contextual shifts in multidimensional experiments occurred 'automatically and preattentively'. Marks (Marks, 1993) found a 'contrastive or adaptationlike' effect when auditory stimuli varied multidimensionally, and assimilative effects when stimuli varied unidimensionally. He argued that these results were consistent with an early process of peripheral adaptation and a later process of response assimilation. Later, the contrastive effect was shown to depend upon stimulus separation (Marks, 1994). Similarly, in considering range effects, Algom and Marks (Algom and Marks, 1996) found binaural loudness integration (as measured by monaural matching) to depend upon stimulus range, once again con-sistent with perceptual rather than response-based shifts.

If an early perceptual encoding process or Wedell's stage of dimensional placement is involved, it would not be surprising to find that all response methods entail some degree of susceptibility to contextual influences. To the extent that scaling methods are able to uncover true sensory or perceptual shifts due to context and that these are worthy of study in their own right, there is merit to each of these procedures. This is a different outlook on contextual effects, which were early on considered to be merely sources of unwanted bias or error in scaling studies, but which have lately become legitimized phenomena for further study.

In Experiments 1 and 2, the magnitude estimation tech-nique with a reference showed a smaller contextual shift than those observed with the other scaling methods. Several explanations of this effect are possible. First, Parducci and Wedell (Parducci and Wedell, 1986) found that as the number of categories in a response scale increased, the magnitude of contextual shifts decreased. Magnitude estimation, as an open-ended continuous numerical scale, could represent an extreme on this continuum of response options. If numbers of response options exerts a minimizing effect on the contrast shift, it is noteworthy that the LMS scale showed strong contextual shifts. This would imply that the LMS may be treated as offering just a few limited response options to the users.

Another explanation for the lower contextual shift in magnitude estimation is the possible stabilizing influence of the reference stimulus given in that procedure. It is widely believed, although rarely demonstrated, that trained panel-ists in applied descriptive analysis techniques, such as the

Texture Profile method (Brandt *et al.*, 1963; Muñoz, 1986) or the Spectrum Descriptive analysis technique (Meilgaard *et al.*, 1991), can be calibrated and anchored to physical examples to stabilize and equate their scale usage. The issue of whether training of industrial product testing panels can eliminate or minimize contextual shifts remains an open question. Further research is needed to determine the extent to which such training and calibration can inoculate sensory judges against the tendency to shift as a function of immediate stimulus context and range. A related issue is the extent to which reference items, such as the modulus in a magnitude estimation method, can prevent or minimize scale shifting. Of course, these practical uses of scaling for product evaluations are very different from physiologically oriented studies such as investigations of PROP tasters responses, aging or sensory deficit studies or clinical investigations. In such studies, precalibration of response ranges, such as the use of a reference modulus in magnitude estimation, would render comparisons among individuals meaningless. A scale such as the LMS might facilitate inter-individual comparisons. However, the current results suggest that as long as those individuals are experiencing different contexts, their data should not be compared.

References

Algom, D. and **Marks, L.E.** (1996) *Range and regression, loudness scales and loudness processing: toward a context bound psychophysics*. J. Exp. Psychol. Hum. Percept. Perf., 16, 706–727.

Borg, G. (1982) *A category scale with ratio properties for intermodal and interindividual comparisons*. In Geissler, H.G. and Petzold (eds), Psychophysical Judgement and the Process of Perception. VEB Deutscher Verlag der Wissenschaften, Berlin, pp. 25–34.

Brandt, M.A., Skinner, E.Z. and **Coleman, J.A.** (1963) *Texture profile method*. J. Food Sci., 28, 404–409.

Diehl, R.L., Elman, J.L. and **McCusker, S.B.** (1978) *Contrast effects in stop consonant identification*. J. Exp. Psych.: Hum. Percept. Perf., 4, 599–609.

Engen, T. and **Levy, N.** (1958) *The influence of context on constant-sum loudness judgements*. Am. J. Psychol., 71, 731–736.

Green, B.G., Shaffer, G.S. and **Gilmore, M.M.** (1993) *Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties*. Chem. Senses, 18, 683–702.

Green, B.G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K. and **Higgins, J.** (1996) *Evaluating the labeled magnitude scale for measuring sensations of taste and smell*. Chem. Senses, 21, 323–334.

Helson, H.H. (1964) Adaptation-level Theory. Harper & Row, New York.

Lawless, H.T. (1983) *Contextual effects in category ratings*. J. Testing Eval., 11, 346–349.

Lawless, H.T. and **Malone, G.J.** (1986) *A comparison of scaling methods: sensitivity, replicates and relative measurement*. J. Sensory Stud., 1, 155–174.

Lawless, H.T., Glatter, S. and **Hohn, C.** (1991) *Context dependent changes in the perception of odor quality*. Chem. Senses, 16, 349–360.

Manis, M. (1967) *Context effects in communication*. J. Person. Soc. Psychol., 5, 326–334.

Marks, L.E. (1992) *The slippery context effect in psychophysics; intensive, extensive and qualitative continua*. Percept. Psychophys., 51, 187–198.

Marks, L.E. (1993) *Contextual processing of multidimensional and unidimensional auditory stimuli*. J. Exp. Psychol. Hum. Percept. Perf., 19, 1–23.

Marks, L.E. (1994) *'Recalibrating' the auditory system: the perception of loudness*. J. Exp. Psychol. Hum. Percept. Perf., 20, 382–396.

Marks, L.E., Borg, G. and **Ljunggren, G.** (1983) *Individual differences in perceived exertion assessed by two new methods*. Percept. Psychophys., 34, 280–288.

Mattes, R.D. and **Lawless, H.T.** (1985) *An adjustment error in optimization of taste intensity*. Appetite, 6, 103–114.

Mellers, B.A. and **Birnbaum, M.H.** (1982) *Loci of contextual effects in judgement*. J. Exp. Psychol. Hum. Percept. Perf., 8, 582–601.

Meilgaard, M., Civille, G.V. and **Carr, B.T.** (1991) Sensory Evaluation Techniques. CRC Press, Boca Raton, FL.

Muñoz, A.M. (1986) *Development and application of texture reference scales*. J. Sensory Stud., 1, 55–83.

Muñoz, A.M. and **Civille, G.V.** (1998) *Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis*. J. Sensory Stud., 13, 57–75.

Parducci, A. (1965) *Category judgement: a range-frequency model*. Psych. Rev., 72, 407–418.

Parducci, A. and **Wedell, D.H.** (1986) *The category effects with rating scales: number of categories, number of stimuli and method of presentation*. J. Exp. Psychol. Hum. Percept. Perf., 12, 496–516.

Poulton, E.C. (1989) Bias in Quantifying Judgements. Lawrence Erlbaum, Hillsdale, NJ.

Rankin, K.M. and **Marks, L.E.** (1991) *Differential context effects in taste perception*. Chem. Senses, 16, 617–629.

Riskey, D.R., Parducci, A. and **Beauchamp, G.K.** (1979) *Effects of context in judgements of sweetness and pleasantness*. Percept. Psychophys., 26, **171**–176.

Schifferstein, H.J.N. (1995) *Contextual shifts in hedonic judgement*. J. Sensory Stud., 381–392.

Schifferstein, H.J.N. and **Frijters, A.E.R.** (1992) *Contextual and sequential effects on judgements of sweetness intensity*. Percept. Psychophys., 52, 243–255.

Simpson, D.D. and **Ostrom, T.M.** (1976) *Contrast effects in impression formation*. J. Personal. Soc. Psychol., 14, 340–350.

Stevens, S.S. (1956) *The direct estimation of sensory magnitudes—loudness*. Am. J. Psychol., 69, 1–25.

Stevens, J.C. and **Marks, L.E.** (1980) *Cross-modality matching functions generated by magnitude estimation*. Percept. Psychophys., 27, 379–389.

Teghtsoonian, R. and **Teghtsoonian, M.** (1978) *Range and regression effects in magnitude scaling*. Percept. Psychophys. 24, 305–314.

Ward, L.M. (1982) *Mixed-modality psychophysical scaling: sequential dependencies and other properties*. Percept. Psychophys. 31, 53–62.

Wedell, D.H. (1990) *Method for determining the locus of context effects in judgements*. In Caverni, J.P., Fabre, J.M. and Gonzalez, M. (eds), Cognitive Biases. Elsevier Science, Amsterdam, pp. 285–304.