Context effects, reliability, and internal consistency of intermodal joint scaling

STEVEN NORDIN
Institute of Environmental Medicine of the Karolinska Institute,
and Stockholm University, Stockholm, Sweden

Context effects, intraindividual variability, and internal consistency of intermodal joint scaling with magnitude estimation ("magnitude matching") were studied by instructing 12 subjects to judge the three pairs of odor intensity, loudness, and brightness on a common scale of perceived intensity as well as to judge odor intensity separately (unimodal magnitude estimation). Significant context effects were found by comparing odor intensity judgments obtained by separate versus intermodal joint scaling as well as across different modalities (loudness vs. brightness) in joint scaling. But no such effects were found for loudness or brightness when compared across modality of joint scaling. Intraindividual variability in the estimates imply about equal reliability in intermodal joint scaling and separate scaling. Good internal consistency was found, indicating that subjects are successful in expressing perceived intensities of different modalities on a common scale.

Information on perceived intensity is required in a number of situations, including clinical settings for diagnosing patients with respect to sensory functions and symptoms, and field settings for investigating either characteristics of the population per se, or characteristics of the environment to which the population is exposed. In such studies, differences between judgments of perceived intensity, between subjects or within the same subject, may or may not have a sensory basis. The differences can arise from two distinct sources. The first is that subjects in the same contextual setting differ with respect to perception and/or manner of expressing perceived intensity, and the second is that the same subject gives different responses due to effects arising from different contextual settings and/or from measurement error variability.

Interindividual Differences in Expressing Perceived Intensity

Interindividual differences in expressing perceived intensity, within a constant contextual setting, may to some extent be eliminated by the method of cross-modality matching (S. S. Stevens, 1959). Intermodal (also referred to as mixed modality) joint scaling is an alternative method to cross-modality matching for taking into consideration differences between subjects in expressing perceived in-

tensity. In intermodal joint scaling, perceived intensities for different modalities are judged on a common scale.

As for cross-modality matching, intermodal joint scaling generates equal-sensation functions, but, in contrast, it can also generate multimodal psychophysical functions on directly comparable psychological scales. Other advantages of intermodal joint scaling, in contrast to cross-modality matching, include (1) freedom from regression effects, (2) application to continua for which continuous stimulus variation is impractical, (3) need for only brief stimulus presentations, which helps control adaptation, and (4) speed (J. C. Stevens & Marks, 1980).

Intermodal joint scaling with magnitude estimation (ME), in which numbers are assigned to perceived intensities in proportional relations, has become a common method of ratio scaling. This combined scaling procedure was originally suggested by J. C. Stevens and Marks (1980), and called *magnitude matching*. However, because magnitude matching can rather easily be confused with the method of cross-modality matching, and because the name refers nonspecifically to either intra- or intermodal scaling, the terms intra- and intermodal joint scaling with ME will be used here.

Intra- or intermodal joint scaling with, for example, ME can be utilized for "calibrating" the psychological scales for individual or group differences in expressing perceived intensity. In intermodal joint scaling, the scale values obtained for one sensory modality, called the reference or standard modality, can be used for calibrating the scale of a second sensory modality, called the target or test modality. The description of such calibration procedures lies outside the scope of this paper, but can be found elsewhere—for example, the normalization procedure (J. C. Stevens, Plantinga, & Cain, 1982), or the master scale procedure (B. Berglund, 1991; B. Berglund & Lind-

vall, 1979). A number of intermodal joint scaling experiments with ME have been performed to study sensory functions in various populations, and in these studies attempts have been made to calibrate the scales for differences in expressing perceived intensity (e.g., Bartoshuk, Rifkin, Marks, & Bars, 1986; B. Berglund & Nordin, 1992; Cometto-Muñiz & Cain, 1982; Marks et al., 1988; J. C. Stevens, Bartoshuk, & Cain, 1984; J. C. Stevens & Cain, 1985, 1986a, 1986b, 1987; J. C. Stevens et al., 1982).

Context Effects

Both the presence (Birnbaum, 1982a, 1982b; Parducci, 1974, 1982; Poulton, 1989) and absence (Gescheider & Hughson, 1991; Hellman, 1976; Zwislocki & Goodman, 1980) of context effects have been demonstrated in separate (unimodal) direct scaling methods. To some extent, context effects have also been studied in intra- and intermodal joint scaling, for example, with respect to sequence (Ward, 1982, 1985). B. Berglund, U. Berglund, and Lindvall (1978) found nearly identical odor intensity scales for both H_2S and n-butanol when scaled intramodally jointly as compared with separately. Zwislocki and Goodman (1980), using auditory stimuli, reported a smaller slope in intramodal joint scaling than in separate scaling.

It has been suggested that subjects compromise between absolute and relative ME in both intramodal (Marks, 1988) and intermodal (Marks, Szczesiul, & Ohlott, 1986) joint scaling. In relative ME, the estimates of one modality change as a function of mean stimulus intensity of the other modality. Marks (1991a) suggested that the size of this change increases with an increase in qualitative difference between the stimuli that are being jointly scaled. Support for this hypothesis has been found for intramodal joint scaling of loudness (Marks, 1991a, 1992, 1993) and for both intra- and intermodal joint scaling of taste and odor intensity (Rankin, 1993a, 1993b; Rankin & Marks, 1991, 1992). Surprisingly, Rankin found context effects even though the taste and odor stimuli were judged as being qualitatively very similar, and independent of orthonasal or retronasal presentation.

Marks's (1991a) hypothesis and findings suggest serious limitations in the application of joint scaling. For reliable results, it may be necessary to choose stimulus intensity ranges for the target and reference modalities that do not differ substantially in perceived intensity. Further support for choosing such ranges lies in the nature of the typical application of joint scaling experiments, in which the target-modality judgments are calibrated (transformed) for scaling behavior by use of the reference-modality judgments. If the target judgments fall far outside the boundaries of the reference judgments, there is not sufficient information on the reference modality to base the calibration of the target judgments. The optimal condition would be to use a reference range that is broader than the target range, thereby "covering" the target range. However, this was not possible in the present experiment, because each modality functioned as both a target and reference

modality. Related to Marks's hypothesis is the question of whether perceived intensity scales for the target modality judged by intermodal joint scaling with qualitatively different reference modalities still differ when steps have been taken to present somewhat similar perceptual ranges for the target and reference modalities. This was one of the objectives of the present study. In addition, odor intensity data obtained by intermodal joint scaling were compared with data from separate (unimodal) scaling with ME (S. S. Stevens, 1975).

Variability

J. C. Stevens and Marks (1980) showed that for data generated by intermodal joint scaling with ME (loudness and brightness), the interindividual variability in absolute magnitude of the estimates and in the exponents were smaller when plotted as equal-sensation functions than when plotted as separate psychophysical functions. The authors explained this finding by referring to the notion that a subject who gives a high exponent for one modality also tends to do so for the other. Marks (1991b) has also shown that differences across individuals in equal-sensation functions generated by intermodal joint scaling greatly exceeded the fluctuations within individuals, which agrees well with results on separate ME (e.g., B. Berglund & Nordin, 1990). In the present study, the intraindividual variability in the magnitude estimates was used as a measure of reliability in comparing separate scaling and intermodal joint scaling of different combinations of target and reference modalities.

Internal Consistency

Several experiments have been studies of the intermethod consistency between separate ME and cross-modality matching, in which predicted exponents for equal-sensation functions generated by separate ME were compared with the same exponents empirically obtained by cross-modality matching (e.g., S. S. Stevens, 1959). In the present study, the consistency within the same scaling method, intermodal joint scaling, was investigated in order to evaluate whether subjects can successfully compare perceived intensities of different qualities (modalities) and express them on a common scale of perceived intensity. This was performed by comparing predicted and empirical equal-sensation functions. The functions were described in terms of exponents and ordinate values corresponding to an invariant abscissa value, used to describe the function's vertical position.

Objectives

Odor intensity was scaled separately with ME, and the three pairs of odor intensity, loudness, and brightness were judged by intermodal joint scaling with ME. Specifically, the purpose was to study (1) context effects on odor intensity scales generated by intermodal joint scaling with loudness and brightness as compared with odor intensity scales generated by separate scaling; (2) context effects on the scales of the target modalities of odor in-

tensity, loudness, and brightness generated by intermodal joint scaling when varying reference modalities; (3) intraindividual variability in magnitude estimates for odor intensity generated by separate scaling and for odor intensity, loudness, and brightness generated by intermodal joint scaling with different reference modalities; and (4) internal consistency in intermodal joint scaling.

METHOD

Subjects

Twelve subjects participated (Subjects A-L; 8 women and 4 men), ranging in age from 20 to 30 years ($M = 23.3$, $SD = 3.4$). According to self-reports, all the subjects had normal sensitivities to odor, sound, and light; none were smokers, had allergies or allergy-like hypersensitivity, or had worked in environments with chemical or physical factors that could have had an effect on their sensory sensitivities.

Stimuli and Apparatus

Odor. Nine concentrations of pyridine ranging from 300 to 2,000 ppb were used (geometric series; step size = 0.103 log units); they were selected in order to include only clearly detectable stimuli, but to exclude unpleasant intensities. Each concentration was repeated eight times and presented in an irregular order in such a way that each was followed by every other concentration once. The procedure led to a series of 73 stimuli with one and the same stimulus concentration presented as the first and the last stimulus, thus, presented nine times. However, the very first magnitude estimate was excluded, leaving eight repetitions for further calculations.

The experiment was performed in an odor laboratory in which air contaminants are kept at low levels by charcoal filtering the intake air. The laboratory essentially consists of a ventilated and air-conditioned test room with an odor-exposure hood connected to a dynamic olfactometer installed in an adjacent room. The olfactometer allows for rapid changes of well-controlled concentrations (B. Berglund, U. Berglund, Johansson, & Lindvall, 1986). The exposure hood has an opening for the subject's nose and mouth and allows for fairly natural breathing conditions. For each odor presentation, the subject took one sniff with mouth closed and then withdrew from the hood. The temperature and relative humidity in the test room were $22° \pm 1°C$ and $40\% \pm 2\%$ RH, respectively.

A base concentration of pyridine was formed by flowing charcoal-filtered air through a glass flask containing liquid pyridine. The base concentration was measured and monitored by an infrared-analyzing instrument (Foxboro-Wilks, Miran 1A) and adjusted continuously to be kept constant during the experimental sessions. The nine pyridine concentrations were obtained with the aid of combinations of steel capillaries of varying lengths and diameters. By means of magnetic valves, the gas flow from the capillaries were injected into the continuous air flow through the exposure hood (100 L/min).

Sound and light. Because somewhat similar perceptual ranges for the target and reference modalities were required for the present study, 8 subjects (who did not participate further in the experiment) were instructed to match loudness and brightness to the perceived odor intensities of the 300 and 2,000 ppb pyridine concentrations. Nine sound pressure levels (SPLs) of pink noise (-3 dB/octave) were selected, ranging from 41 to 65 dB with 3-dB intervals. The pink noise was generated by a random-noise generator (Brüel & Kjaer, Type 1402), recorded on a four-channel cassette recorder (Tascam, Portastudio 244), and attenuated (Danbridge, DA3T/D) to the nine SPLs. The presentation was binaural via earphones (AKG, K109S) and calibrated with the aid of an artificial ear (Brüel & Kjaer, Type 4153).

Nine luminance levels of white light were used, ranging from 10.0 to 2,200 cd/m² (geometric series; step size = 0.293 log units). The light was generated by a white incandescent lamp placed in a box with an internal white color. The dimensions were 0.40 m high × 0.28 m wide × 0.14 m deep, with a 160-mm-diam opening facing the subject. No direct radiation from the lamp reached the subject, who sat at a distance of approximately 0.8 m from the illuminated area. The light was attenuated by a transformer to the nine selected luminance levels, and calibrated with the aid of a luminance meter (Brüel & Kjaer, Type 1100).

The stimulus duration for all three modalities was fixed at 5 sec, and the interstimulus interval (ISI) with respect to offset and onset was approximately 15 sec. However, because the duration of one sniff was approximately only 1 sec, the actual ISI was 19 sec between the sniffs, and 17 sec between sniffs and sounds as well as between sniffs and lights. Each of the nine luminances and SPLs was presented eight times (except for one level, repeated nine times) and presented in irregular orders in principally the same manner as the odor stimulus series. Thus, a series of 73 sound stimuli and a series of 73 light stimuli were obtained (the very first estimate was excluded from further calculations).

Conditions and sessions. The experiment was divided into four conditions for each subject: Condition 1, in which odor intensity was scaled separately with ME; Condition 2, in which odor intensity and loudness were scaled jointly with ME; Condition 3, in which odor intensity and brightness were scaled jointly with ME; and Condition 4, in which loudness and brightness were scaled jointly with ME. The combined sets were arranged so that, for example, the first odor stimulus in the odor series was followed by the first sound stimulus in the sound series, which, in turn, was followed by the second odor stimulus, and so on.

The four conditions required a total of 14 sessions. The odor stimulus series (Condition 1) was divided into two sessions, and each of the three combined sets (Conditions 2–4) was divided into four sessions. To accomplish this, each session comprised 36 or 37 stimuli. Each subject had a unique order of the 14 sessions, which were randomized so that two sessions of the same condition never succeeded each other.

Procedure

Each subject's participation required 2 consecutive experimental days. A prerequisite for participation was that the subject was able to perform a separate scaling task with ME (S. S. Stevens, 1975). To accomplish this, the potential subject started the first experimental day by judging visual line length with free-number ME. The subject was to assign an appropriate number to the first length; the second length was then to be compared with the first one (the third with the second, etc.) in proportion. Seven black lines (5, 9, 15, 27, 45, 85, and 150 mm, with widths of 0.5 mm) were presented on a visual display terminal (175 × 115 mm). The seven line lengths were presented five times each in a randomized order. All judgments were given orally by the subject via a microphone. The judgments of visual line length were followed by a 5-min period, during which time it was decided whether the subject would participate further. Thereafter, the subject was given instructions and took part in a training session while being preconditioned to air purified by charcoal filters (15 min). This 15-min period also initiated the second experimental day. For both days, the preconditioning period was followed by seven sessions separated by 5-min pauses, except for the third and fourth sessions, which were separated by a 15-min pause.

For Condition 1, the subjects were instructed to judge the perceived odor intensities in the same manner as they judged the visual line lengths (free-number ME). Similar instructions were given for Conditions 2–4, except that for these conditions the subjects were instructed to judge the perceived intensities for the two modalities on a common scale of perceived intensity. For example, an odor

and sound perceived to have the same intensity should be assigned the same number; an odor perceived as twice as intense as the loudness of a sound should be assigned a number twice as large as the number assigned to the loudness. Thus, for Conditions 2-4, explicit instructions were given to the subjects to (1) judge the perceived intensity for a sensory experience in relation to the previously experienced one, and (2) use the same scale for all four conditions over the 14 sessions.

RESULTS

General Data Treatment

The intraindividual variability in response ranges for the various modalities and conditions was determined for each subject, for each of the seven data sets: (1) odor intensity scaled separately, (2) odor intensity scaled intermodally and jointly with loudness, (3) odor intensity scaled jointly with brightness, (4) loudness scaled jointly with odor intensity, (5) loudness scaled jointly with brightness, (6) brightness scaled jointly with odor intensity, and (7) brightness scaled jointly with loudness. Figure 1 presents the individual response ranges, including those for visual line length, for Subjects A-L. The two endpoints of the bars represent the ranges corresponding to the stimuli with the highest and lowest geometric mean estimate. The results suggest that the response ranges, within subjects, for the different modalities and conditions, do differ somewhat (which permits an internal-consistency analysis). Over subjects, the ranges for loudness and brightness are similar, whereas for odor intensity, the ranges seem to be slightly lower.

Individual and group psychophysical power functions for the seven data sets and for visual line length were obtained by averaging (geometric means) the estimates for each stimulus intensity over stimulus repetitions (for the group, geometric mean over subjects as well). Regression lines were fitted to the averaged data (in double log coordinates) by the method of least squares, and individual and group exponents, multiplicative constants, and Pearson correlation coefficients were determined. Large correlation coefficients, averaging .97 (range = .75-.99), imply that the functions can adequately be described as power functions according to Equation 1:

$$\Psi = c\phi^n,\tag{1}$$

where Ψ represents the perceived intensity, ϕ the stimulus intensity, n the exponent, and c the multiplicative constant (cf. S. S. Stevens, 1975). Individual and group grand means of magnitude estimates were obtained for the seven data sets for describing the vertical positions of the perceptual scales.

The exponent for odor intensity ($M = 0.97$; range = 0.26-2.00) and loudness ($M = 0.79$; range = 0.58-1.04) are rather large. This may be due to a range effect (Poulton, 1989), because relatively small ranges were used to avoid unpleasant (odor) intensities. The results support previous findings of a large interindividual variability in exponent size for pyridine odor intensity (e.g., B. Berglund, Högman, & Johansson, 1988).

Ability to Scale Separately

Results from several studies have shown that the exponent of the psychophysical power function for visual line length is approximately unity (Hellman & Meiselman, 1988; S. S. Stevens & Guirao, 1963). Therefore, the exponent for line length and the goodness of fit were used in the present study as a reference for evaluating whether the subject was likely to be able to perform a separate scaling task, and would thereby be allowed further participation. The criteria for acceptable exponents and goodnesses of fit were arbitrarily set at 0.7-1.3 and above 0.95, respectively. All 12 subjects met the criterion for the exponent (range = 0.82-1.30) and goodness of fit (range = 0.98-0.99), and were allowed further participation.

Context Effects

Individual and group exponents and grand means for separate scaling of odor intensity are plotted in Figure 2 against corresponding values for odor intensity scaled jointly with loudness and brightness. Eleven and 10 of the 12 subjects have smaller exponents for odor intensity scaled jointly with loudness and brightness, respectively, than when scaled separately. A difference in individual grand means is evident as well; all the subjects have larger grand means for odor intensity when scaled jointly with loudness and brightness than when scaled separately.

In Figures 3 and 4, individual and group exponents and grand means for the target modalities of odor intensity, loudness, and brightness scaled jointly with one reference modality are plotted against corresponding values for the same target modalities scaled jointly with another modality. The figures show that, whereas larger exponents and grand means are found for odor intensity scaled jointly with brightness as compared with loudness, no such context effects are found for loudness and brightness.

The differences in individual odor intensity exponents and grand means (log units) for the three scaling contexts were tested statistically by one-way analyses of variance with repeated measures, showing significantly different exponents [$F(2,11) = 11.21, p < .001$] and grand means [$F(2,11) = 42.12, p < .0001$]. Post hoc paired t tests revealed significantly lower exponents for odor intensity scaled jointly with loudness [$t(11) = 4.26, p < .01$] and brightness [$t(11) = 2.69, p < .05$] than when separately scaled, as well as larger exponents for odor intensity scaled jointly with brightness than jointly with loudness [$t(11) = 2.53, p < .05$]. In addition, significantly larger grand means were found for odor intensity scaled jointly with loudness [$t(11) = 5.69, p < .0001$] and brightness [$t(11) = 7.87, p < .0001$] than when separately scaled, as well as larger grand means for odor intensity scaled jointly with brightness than jointly with loudness [$t(11) = 3.34, p < .01$].

A priori paired t tests were performed on the data for the target modalities of loudness and brightness when varying the reference modality. These analyses showed no significant differences for loudness exponents [$t(11) = 0.35$], brightness exponents [$t(11) = 1.17$], loudness grand means [$t(11) = 2.09$], or brightness grand means [$t(11) = 1.11$].
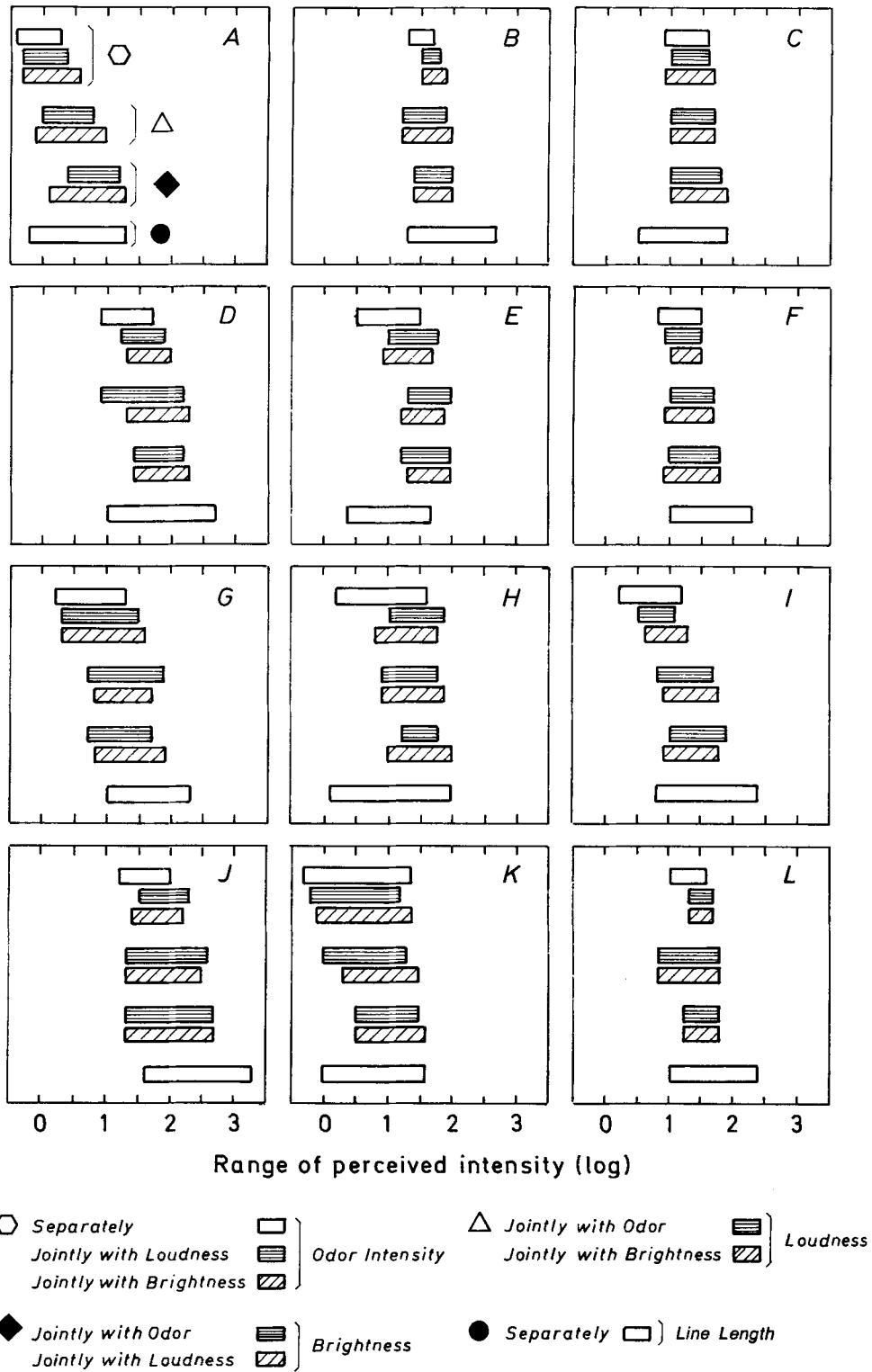
Figure 1. Response ranges for Subjects A–L for odor intensity scaled separately, intermodally and jointly with loudness, and jointly with brightness, for loudness scaled jointly with odor intensity and jointly with brightness, for brightness scaled jointly with odor intensity and jointly with loudness, and for visual line length scaled separately.
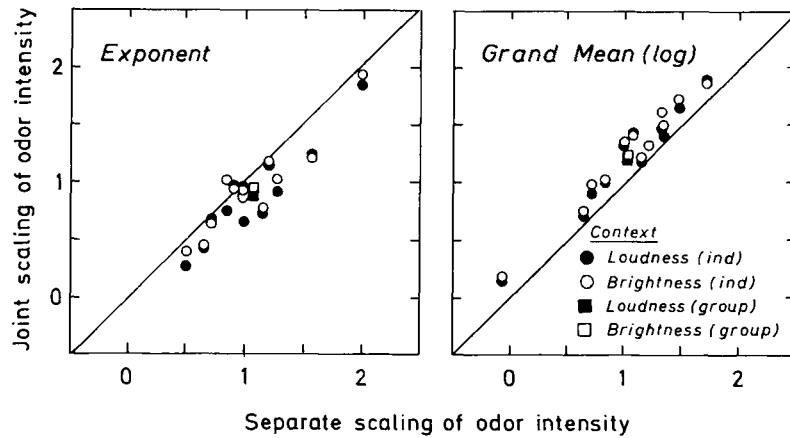
Figure 2. Exponents for individual and group psychophysical power functions and individual and group means (log units) for odor intensity judgments scaled separately and intermodally and jointly with loudness and brightness. The principal diagonals represent identity.

Variability

Intraindividual coefficients of variation (defined as SD/AM) in the perceived intensity estimates were calculated for each of the seven data sets, for each stimulus intensity, and for each subject. The average (arithmetic mean) coefficients over subjects for each stimulus intensity are presented in Figure 5 as a function of average (geometric mean) perceived intensity. A marked feature is a decrease in relative variability with an increase in perceived intensity.

In order to provide a direct comparison between the seven data sets, linear regression lines were fitted to the mean coefficients for each data set. On the basis of these regression lines, mean coefficients corresponding to the perceived intensity of 10 (1.0 in log units) were calculated, permitting interpolation in the seven data sets. For odor intensity, the coefficients are 0.60 for separate scaling, 0.66 for intermodal joint scaling with loudness, and 0.59 for

joint scaling with brightness. For loudness, the coefficients are 0.47 for joint scaling with odor intensity and 0.45 for joint scaling with brightness. And for brightness, the coefficients are 0.39 for joint scaling with odor intensity and 0.37 for joint scaling with loudness. Referring to the calculated coefficients as well as to the general trends of the data sets, the largest variability was found for odor intensity, followed by loudness and brightness, in that order. However, there seems to be no difference in variability in odor intensity estimates scaled separately and scaled jointly. In addition, the variability in the target modality estimates seems to be invariant, regardless of the reference modality.

Internal Consistency

Internal consistency in intermodal joint scaling was studied by comparing predicted and empirical individual and group equal-sensation power functions for (1) odor concentration and SPL, (2) odor concentration and lu-
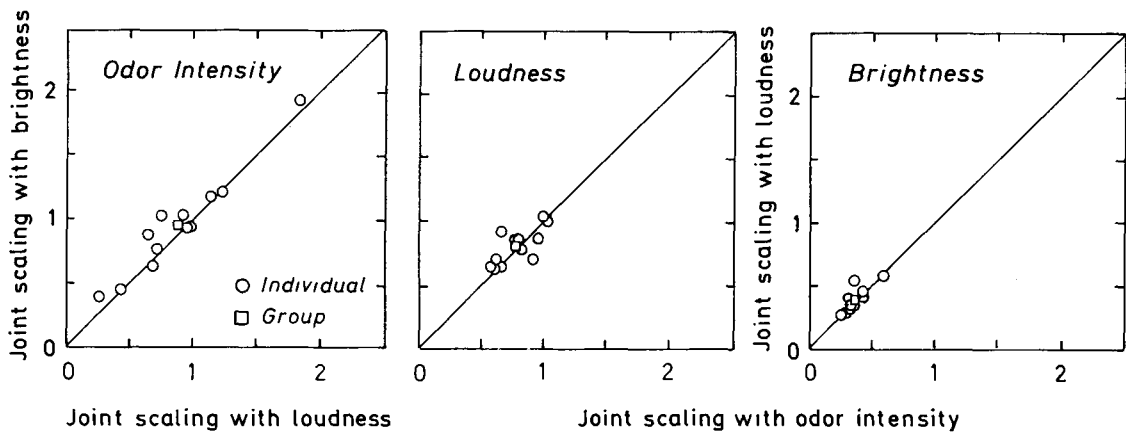

Figure 3. Exponents for individual and group psychophysical power functions for odor intensity, loudness, and brightness scaled intermodally and jointly with two different modalities. The principal diagonals represent identity.
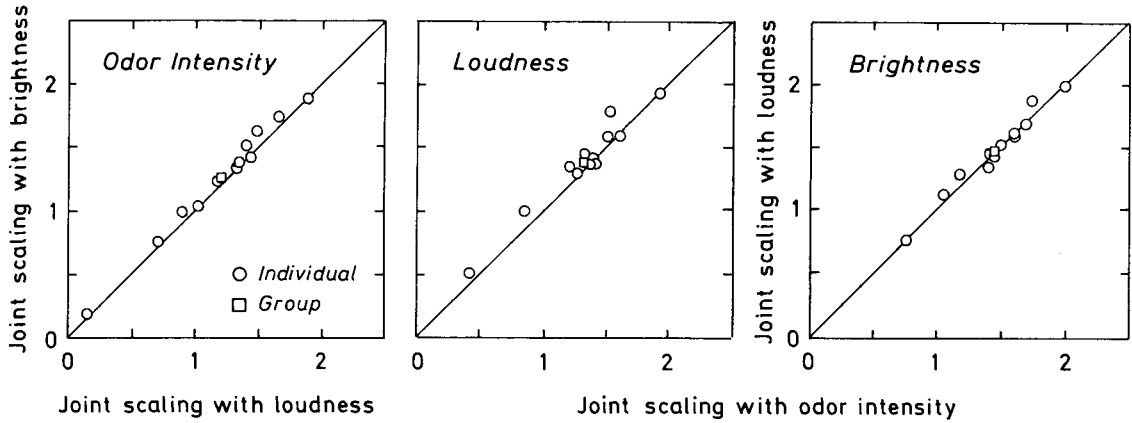
Figure 4. Grand means (log units) for individual and group judgments of odor intensity, loudness, and brightness scaled intermodally and jointly with two different modalities. The principal diagonals represent identity.

minance level, and (3) SPL and luminance level. The relations between perceived intensity and stimulus intensity were found to be adequately described as power functions (Equation 1), so the functions obtained by intermodal joint scaling can, in double log units, be written as

$$y_{i,j} = b_{i,j}x_i + a_{i,j}, \qquad (2)$$

where index i represents the modality being scaled (target), and index j represents the modality with which the target i is jointly scaled (reference).

Because the subjects were instructed to judge the perceived intensities for the two modalities on a common scale, $y_{i,j}$ can be set equal to $y_{j,i}$, leading to three empirical equal-sensation functions of the form

$$x_i = k_{i,j}x_j + c_{i,j}, \qquad (3)$$

where $k_{i,j} = b_{j,i}/b_{i,j}$ and $c_{i,j} = (a_{j,i}-a_{i,j})/b_{i,j}$. The vertical position of the empirical equal-sensation function is defined as the ordinate value corresponding to the invariant abscissa value of the middle stimulus intensity (log

values of 775 ppb and 148 cd/m², and the linear value of 53 dB) by use of Equation 3.

In addition to these empirical functions, it is possible to derive, for example, the equal-sensation relation between odor and loudness, predicted from the relation between odor and brightness and that between brightness and loudness. The predicted equal-sensation functions are:

$$x_1 = (1/k_2k_3)x_3 - (k_2c_3+c_2)/k_2k_3 \qquad (4)$$

$$x_2 = (1/k_1k_2)x_1 - (k_1c_2+c_1)/k_1k_2 \qquad (5)$$

$$x_3 = (1/k_3k_1)x_2 - (k_3c_1+c_3)/k_3k_1, \qquad (6)$$

where index 1 refers to odor intensity, index 2 refers to loudness, and index 3 refers to brightness. Thus, the predicted exponents k'_1, k'_2, and k'_3 (cf. Equation 3) are $1/k_2k_3$, $1/k_1k_2$, and $1/k_3k_1$, respectively, whereas the predicted vertical positions were obtained from Equations 4-6 by use of the given invariant abscissa values.

Empirical exponents for odor concentration, luminance level, and SPL as equal-sensation functions of SPL, odor
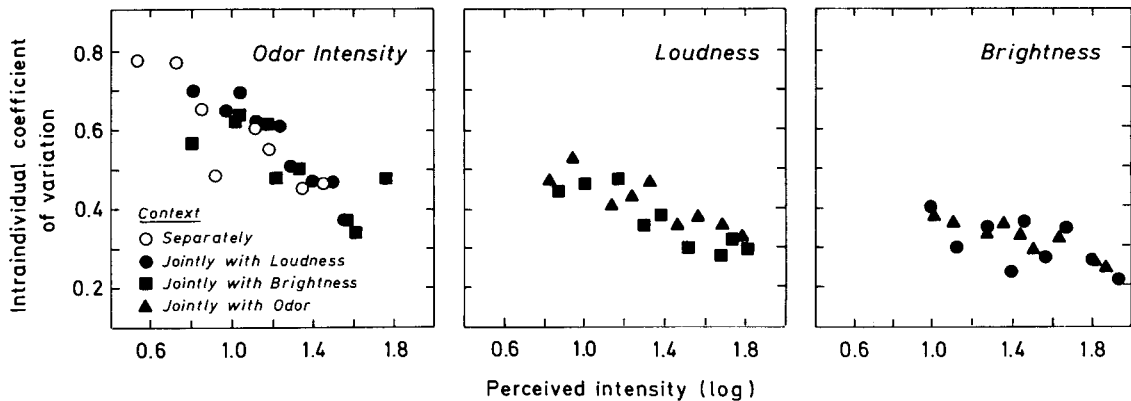

Figure 5. Intraindividual coefficients of variation for each stimulus intensity averaged over subjects as functions of corresponding group means for perceived intensity (log units). The coefficients are presented for odor intensity scaled separately, intermodally and jointly with loudness and jointly with brightness, for loudness scaled jointly with odor intensity and jointly with brightness, and for brightness scaled jointly with odor intensity and jointly with loudness.
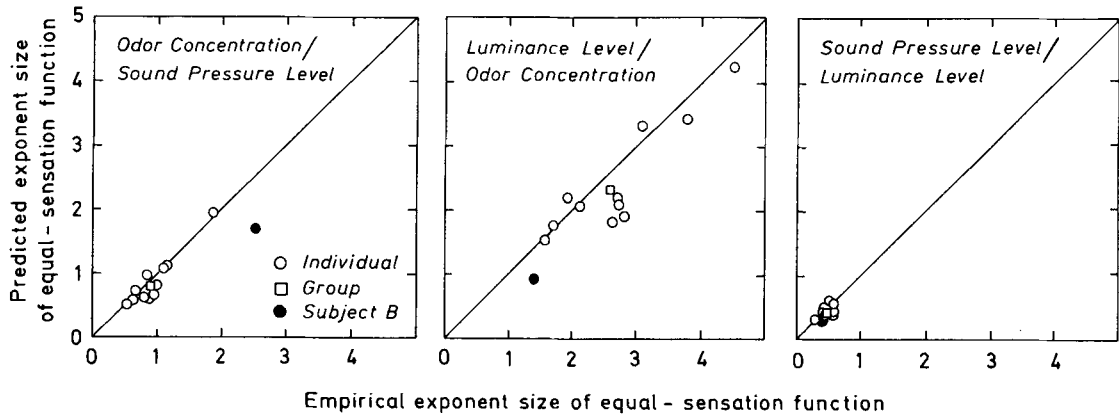
Figure 6. Empirical and predicted individual and group exponents for odor concentration as an equal-sensation power function of sound pressure level, for luminance level as a function of odor concentration, and for sound pressure level as a function of luminance level. The diagonals represent identity.

concentration, and luminance level, respectively, are plotted against corresponding predicted values in Figure 6. The vertical positions are plotted in the same manner in Figure 7. The figures show that for all the subjects, except 1 (Subject B), and for the group as a whole, the predicted values lie rather close to the empirical. It should be noted that the comparisons for the three sets of equal-sensation functions are not independent. However, taken together, they contribute to a more complete description of the internal consistency than would one single set.

DISCUSSION

Context Effects

The present results demonstrate that odor intensity scales generated by intermodal joint scaling differ significantly from odor intensity scales generated by separate scaling. In general, lower individual exponents of psychophysical power functions and larger individual grand means of odor intensity estimates were obtained in inter-

modal joint scaling than in separate scaling. Hence, scales generated by intermodal joint scaling may not reliably be generalized to a separate scaling context. The shift in exponent supports the finding of Zwislocki and Goodman (1980), showing a smaller exponent in intramodal joint scaling than in separate scaling. These authors suggested that the exponent decreases as the complexity of the task increases. The larger grand means may be due to an assimilation effect (e.g., DeCarlo & Cross, 1990; Melamed, 1970)—the response range for odor intensity is, in general, slightly lower than for loudness and brightness (cf. Figure 1). Thus, an "assimilation" of the response for odor intensity toward the preceding loudnesses and brightnesses may have taken place, resulting in larger magnitude estimates for odor intensities scaled jointly than for those scaled separately. Results in accordance with these have also been demonstrated in intramodal joint scaling for loudness (Marks, 1993). Empirical findings on separate scaling suggest that assimilation effects are larger in ratio ME than in absolute ME (DeCarlo & Cross, 1990;
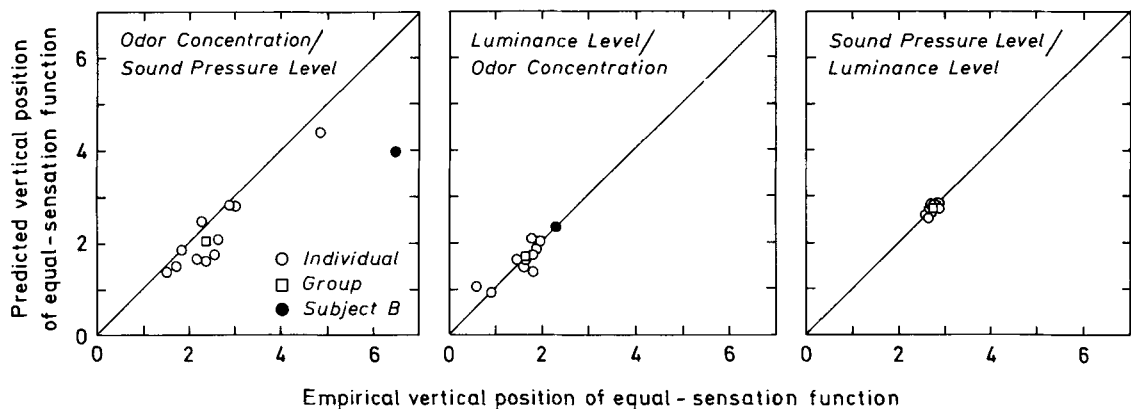

Figure 7. Empirical and predicted individual and group measures of vertical position (see text) for odor concentration as an equal-sensation power function of sound pressure level, for luminance level as a function of odor concentration, and for sound pressure level as a function of luminance level. The diagonals represent identity.

Ward, 1987). However, the extent to which these results can be generalized to joint scaling procedures is not known.

The results further suggest that perceived intensity scales for loudness and brightness generated by intermodal joint scaling do not differ due to the use of different reference modalities, as long as somewhat similar perceptual ranges for the target and reference stimuli are used. For odor intensity, on the other hand, context effects do seem to be present. Using odor and taste as modalities (judged as qualitatively very similar), Rankin (1993a, 1993b) demonstrated context effects in intermodal joint scaling when presenting different mean intensities for the two modalities. As demonstrated by the present findings, context effects exist even when somewhat similar mean intensities are presented. Thus, the exponents and grand means for odor intensity scaled jointly with brightness were significantly larger than those for odor intensity scaled jointly with loudness. Phenomenologically, odors may be less distinct than sounds and lights, in turn, making odor intensity judgments more susceptible to context effects than loudness and brightness.

In applied joint scaling experiments, in which the scale values for the target modality are calibrated for scaling behavior by use of the reference judgments, the reference-modality range would need to be broader than the target range. Such ranges may easily be achieved from a subgroup of subjects, for example, by applying Borg's (1982) category-ratio scale.

Variability

According to the intraindividual coefficients of variation in the magnitude estimates for odor intensity scaled separately and jointly with loudness and brightness, the reliability seems to be about the same for intermodal joint scaling and separate scaling. In addition, the variability in the target-modality estimates appears to be invariant, no matter with what reference modality the target modality is jointly scaled. Variability is slightly smaller in the brightness estimates than in the loudness and odor intensity estimates. Perhaps a slight change in color in the incandescent lamp, accompanying the voltage adjustment, made the visual stimulus multidimensional, thereby providing the subject with a "clue" to intensity. The decrease in variability with an increase in perceived intensity suggests that the number of repetitions of magnitude estimates required for obtaining reliable average estimates is relatively large for weak stimulus intensities. The size of the overall intraindividual variability for odor intensity (about 0.5) and loudness (about 0.4) are the same as previously found for odor intensity (intermodal joint scaling; B. Berglund & Nordin, 1992) and loudness (separate scaling; B. Berglund & Nordin, 1990).

Internal Consistency

The good agreement between predicted and empirical exponents of equal-sensation power functions, as well as between predicted and empirical vertical positions of the

scales, favor the notion that not only a group as a whole, but also individual subjects, in general, are successful in comparing perceived intensities of different qualities (modalities) and in expressing them on a common scale of perceived intensity. These abilities are basic to the application of intermodal joint scaling.

REFERENCES

BARTOSHUK, L. M., RIFKIN, B., MARKS, L. E., & BARS, P. (1986). Taste and aging. *Journal of Gerontology*, **41**, 140-162.

BERGLUND, B. (1991). Quality assurance in environmental psychophysics. In S. J. Bolanowski, Jr. & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitudes: In honor of the memory of S. S. Stevens* (pp. 140-162). Hillsdale, NJ: Erlbaum.

BERGLUND, B., BERGLUND, U., JOHANSSON, I., & LINDVALL, T. (1986). Research equipment for air quality studies of nonindustrial environments. *Environment International*, **12**, 189-194.

BERGLUND, B., BERGLUND, U., & LINDVALL, T. (1978). Separate and joint scaling of perceived odor intensity of *n*-butanol and hydrogen sulfide. *Perception & Psychophysics*, **23**, 313-320.

BERGLUND, B., HÖGMAN, L., & JOHANSSON, I. (1988). *Reliability of odor measurements near threshold* (Tech. Rep. No. 682). Stockholm: Reports from the Department of Psychology, University of Stockholm.

BERGLUND, B., & LINDVALL, T. (1979). Olfactory evaluation of indoor air quality. In P. O. Fanger & O. Valbjørn (Eds.), *Indoor climate: Effects on human comfort, performance, and health* (pp. 141-157). Copenhagen, Denmark: Danish Building Research Institute.

BERGLUND, B., & NORDIN, S. (1990). Utilizing individual differences in loudness measurement. In F. Müller (Ed.), *Proceedings of the Sixth Annual Meeting of the International Society for Psychophysics: Fechner Day '90* (pp. 117-122). Würzburg: Institut für Psykologie, Universität Würzburg.

BERGLUND, B., & NORDIN, S. (1992). Detectability and perceived intensity for formaldehyde in smokers and nonsmokers. *Chemical Senses*, **17**, 291-306.

BIRNBAUM, M. H. (1982a). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attributes and psychophysical measurement* (pp. 401-485). Hillsdale, NJ: Erlbaum.

BIRNBAUM, M. H. (1982b). Problems with so-called "direct" scaling. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewic (Eds.), *Selected sensory methods: Problems and approaches to hedonics* (pp. 34-48). Philadelphia, PA: American Society for Testing and Materials.

BORG, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geisser & P. Petzold (Eds.), *Psychophysical judgment and the process of perception* (pp. 25-34). Berlin: VEB Deutscher Verlag der Wissenschaften.

COMETTO-MUÑIZ, J. E., & CAIN, W. S. (1982). Perception of nasal pungency in smokers and nonsmokers. *Physiology & Behavior*, **29**, 727-731.

DeCARLO, L. T., & CROSS, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, **119**, 375-396.

GESCHEIDER, G. A., & HUGHSON, B. A. (1991). Stimulus context and absolute magnitude estimation: A study of individual differences. *Perception & Psychophysics*, **50**, 45-57.

HELLMAN, R. P. (1976). Growth of loudness at 1000 and 3000 Hz. *Journal of the Acoustical Society of America*, **60**, 672-679.

HELLMAN, R. P., & MEISELMAN, C. H. (1988). Prediction of individual loudness exponents from cross-modality matching. *Journal of Speech & Hearing Research*, **31**, 605-615.

MARKS, L. E. (1988). Magnitude estimation and sensory matching. *Perception & Psychophysics*, **43**, 511-525.

MARKS, L. E. (1991a). The dynamics of ratio scaling. In S. J. Bolanowski, Jr. & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitudes: In honor of the memory of S. S. Stevens* (pp. 27-42). Hillsdale, NJ: Erlbaum.

MARKS, L. E. (1991b). Reliability of magnitude matching. *Perception & Psychophysics*, **49**, 31-37.

MARKS, L. E. (1992). The contingency of perceptual processing: Context modifies equal-loudness relations. *Psychological Science*, **3**, 285-291.

MARKS, L. E. (1993). Contextual processing of multidimensional and unidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 227-249.

MARKS, L. E., STEVENS, J. C., BARTOSHUK, L. M., GENT, J. F., RIFKIN, B., & STONE, V. K. (1988). Magnitude-matching: The measurement of taste and smell. *Chemical Senses*, **13**, 63-87.

MARKS, L. E., SZCZESIUL, R., & OHLOTT, P. (1986). On the crossmodal perception of intensity. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 517-534.

MELAMED, L. E. (1970). The role of response processes in the formation of cross-modality assimilation effects. *Perception & Psychophysics*, **8**, 185-188.

PARDUCCI, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Psychophysical judgment and measurement* (Vol. 2, pp. 127-141). New York: Academic Press.

PARDUCCI, A. (1982). Category ratings: Still more contextual effects. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 89-105). Hillsdale, NJ: Erlbaum.

POULTON, E. C. (1989). *Bias in quantifying judgments*. London: Erlbaum.

RANKIN, K. M. (1993a). *Mechanisms underlying perceptual changes in taste and smell intensities*. (Tech. Rep. No. 765). Stockholm: Reports from the Department of Psychology, University of Stockholm.

RANKIN, K. M. (1993b). *Orthonasal vs. retronasal perception of odour: Role in differential effects of context on taste and smell*. (Tech. Rep. No. 766). Stockholm: Reports from the Department of Psychology, University of Stockholm.

RANKIN, K. M., & MARKS, L. E. (1991). Differential context effects in taste perception. *Chemical Senses*, **16**, 617-629.

RANKIN, K. M., & MARKS, L. E. (1992). Effects of context on sweet and bitter tastes: Unrelated to sensitivity to PROP (6-*n*-propylthiouracil). *Perception & Psychophysics*, **52**, 479-486.

STEVENS, J. C., BARTOSHUK, L. M., & CAIN, W. S. (1984). Chemical senses and aging: Taste versus smell. *Chemical Senses*, **9**, 167-179.

STEVENS, J. C., & CAIN, W. S. (1985). Age-related deficiency in the perceived strength of six odorants. *Chemical Senses*, **10**, 517-529.

STEVENS, J. C., & CAIN, W. S. (1986a). Aging and the perception of nasal irritation. *Physiology & Behavior*, **37**, 323-328.

STEVENS, J. C., & CAIN, W. S. (1986b). Smelling via the mouth: Effect of aging. *Perception & Psychophysics*, **40**, 142-146.

STEVENS, J. C., & CAIN, W. S. (1987). Old-age deficits in the sense of smell as gauged by thresholds, magnitude matching, and odor identification. *Psychology & Aging*, **2**, 36-42.

STEVENS, J. C., & MARKS, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics*, **27**, 379-389.

STEVENS, J. C., PLANTINGA, A., & CAIN, W. S. (1982). Reduction of odor and nasal pungency associated with aging. *Neurobiology of Aging*, **3**, 125-132.

STEVENS, S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, **57**, 201-209.

STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social aspects*. New York: Wiley.

STEVENS, S. S., & GUIRAO, M. (1963). Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, **66**, 177-186.

WARD, L. M. (1982). Mixed-modality psychophysical scaling: Sequential dependencies and other properties. *Perception & Psychophysics*, **31**, 53-62.

WARD, L. M. (1985). Mixed-modality psychophysical scaling: Inter- and intramodality sequential dependencies as a function of lag. *Perception & Psychophysics*, **38**, 512-522.

WARD, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 216-227.

ZWISLOCKI, J. J., & GOODMAN, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, **28**, 28-38.