Perception & Psychophysics
1996, 58 (5), 793-801

Intensity resolution and subjective magnitude in psychophysical scaling

LAWRENCE M. WARD, JULIET ARMSTRONG, and NARLY GOLESTANI
University of British Columbia, Vancouver, British Columbia, Canada

Several successful theories of psychophysical judgment imply that exponents of power functions in scaling tasks should covary with measures of intensity resolution such as d' in the same tasks, whereas the prevailing metatheory of ideal psychophysical scaling asserts the independence of the two. In a direct test of this relationship, three prominent psychophysical scaling paradigms were studied: category judgment without an identification function, absolute magnitude estimation, and cross-modality matching with light intensity as the response continuum. Separate groups of subjects for each scaling paradigm made repeated judgments of the loudnesses of the pure tones that constituted each of two stimulus ensembles. The narrow- and wide-range ensembles shared six identical stimulus intensities in the middle of each set. Intensity resolution, as measured by d'-like distances, of these physically identical stimuli was significantly worse for the wide-range set for all three methods. Exponents of power functions fitted to geometric mean responses, and in magnitude estimation and cross-modality matching the geometric mean responses themselves, were also significantly smaller in the wide-range condition. The variation of power function exponents, and of psychophysical scale values, for stimulus intensities that were identical in the two stimulus sets with the intensities of other members of the ensembles is inconsistent with the metatheory on which modern psychophysical scaling practice is based, although it is consistent with other useful approaches to measurement of psychological magnitudes.

Several authors have argued that all judgment of stimulus intensity has a categorical basis and that scaling methods are no different than other absolute judgment paradigms, such as absolute identification (also called one-interval forced-choice). Braida and Durlach (e.g., 1972, 1988; Durlach & Braida, 1969) have carefully developed, within a Thurstonian or signal detection theory framework, a theory of *auditory intensity resolution* (the ability to distinguish among sounds on the basis of intensity) that they argued underlies all kinds of judgments made of auditory intensity. Intensity resolution is measured in units of d' and is thus closely related to the psychophysical concept of intensity discrimination (this relationship will be discussed in the Discussion section). In this theory, d' depends directly on the physical separation between adjacent stimuli and inversely on two types of variance (or noise) terms—that arising from encoding of stimuli by the auditory system and that arising from memory and judgment factors. Gravetter and Lockhead (1973) developed a similar model with a slightly different partitioning of response variability into stimulus variability and criterial variability. They also argued that "observers judge all intensive stimuli based on the model proposed here" (p. 215). They asserted that among the implications of the model is that the exponent of the psychophysical power function obtained by direct psychophysical scaling techniques (e.g., magnitude estimation, cross-modality matching; see S. S. Stevens, 1975) should be smaller for less resolvable stimuli. The major purpose of the present experiments was a direct test of this assertion.

Ward (1979) also proposed a theory in which judgments made in direct psychophysical scaling techniques are based on categorization of intensive stimuli. In that theory, fuzzy (in the sense of fuzzy set theory; see Zadeh, 1965) psychological representations of stimulus intensities are labeled with the names of fuzzy categories (in a biased way) which are mapped to fuzzy response categories of which satisfactory exemplars are chosen as overt responses. The theory accounts for the sequential dependencies that are ubiquitous in such judgments, but its prediction that responses should depend on the fuzzy categories available, which should in turn depend on intensity resolution, has never been directly tested. The present paper also addresses this prediction.

Regardless of its origin, a close relationship between intensity resolution and sensory magnitude involves two quantities that are not supposed to be related according to the generally accepted metatheories of direct scaling (see, e.g., Gescheider, 1985, for an exposition). The idea that one can obtain the "true" value of stimuli on a scale of subjective magnitude if one simply samples them enough, and that these "true" scale values do not depend on such factors as the amplitude range of the stimuli judged (except for "biases"; cf. Poulton, 1989), also pervades scaling practice, although it has been challenged in various ways throughout its history (e.g., by Algom & Marks, 1990).

More formally, consider a Thurstonian framework for absolute intensity judgment. Several different stimulus in-

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to L.M.W. We thank Bill Turkel for early work on the project and Larry Marks, Ken Norwich, and an anonymous reviewer for their thoughtful reviews. Address reprint requests and comments to L. M. Ward, Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (e-mail: lward@cortex.psych.ubc.ca).

tensities are presented several times for judgment. Each stimulus presentation gives rise to a value on a decision axis that represents a summary of its sensory/perceptual effects. Different presentations of the same stimulus intensity give rise to different values on the decision axis; these values are normally distributed about a mean value, m_i, with standard deviation s_i for stimulus i. In an intensity resolution task—for example, one- or two-interval forced choice—with at least two such stimulus intensities, subjects are assumed to determine a response on each stimulus presentation by comparing the current value of the decision variable (or a difference of two values in the two-interval task) to the values on the same dimension (or difference axis) of one or more criteria, c_{ij}. The goal of the experimenter in an intensity resolution task is to estimate the distance between the means of the stimulus dispersions in units of their mutual standard deviations (d' in signal detection theory) or equivalently to estimate the difference in stimulus intensity that will give rise to a particular value of d' (a jnd-like measure). Using the signal detection theory formulation, the goal is to estimate

$$d'_{ij} = \left(m_j - m_i\right) / s_{ij}. \tag{1}$$

Thurstone (e.g., 1927) recommended creating a scale for the stimulus attribute in question by setting $v_1 = 0$, $v_2 = d'_{12}$, $v_3 = v_2 + d'_{23}$, etc., where v_i represents the scale value of the mean sensory/perceptual effect of stimulus i (m_i). This procedure differs profoundly from direct scaling methods, in which it is assumed that observers can assign numbers (or intensities on another stimulus continuum) proportional to the values experienced on the decision axis so that a measure of the central tendency of the distribution of responses to the several presentations of a given stimulus intensity, say r_i, is proportional to m_i (i.e., $r_i = k\, m_i$). The measure r_i is then taken as the scale value to be assigned to stimulus intensity i. The task of the experimenter in direct scaling is to obtain as reliable a value for r_i as possible and thus locate m_i as accurately as possible relative to the effects of other stimulus intensities. Since a 100 $(1-\alpha)$ % confidence interval for r_i is

$$r_i \pm t_{\alpha/2}\left(s_i / \sqrt{n_i}\right), \tag{2}$$

it is clear that this can be accomplished by increasing n_i (the number of presentations of stimulus i) until the desired accuracy is reached. The crucial observation to make is that the estimate of r_i is assumed not to depend in any way on which other stimuli are being estimated at the same time, or on how close any other stimuli may be to stimulus i. In theory any two stimuli with different intensities give rise to sensory/perceptual effect distributions with different means and the scale values of those stimuli can be determined to the desired degree of accuracy (e.g., a degree that will produce significantly different scale values for the stimuli) simply by taking a sufficient number of observations of each. Notice that such a strategy has no effect whatsoever on the value of d'_{ij} estimated from Equation 1, except to make its error of estimate smaller as well. Similarly, none of the variables that affect d'_{ij}, such as the distances between the stimulus intensities or the variability of the distributions of their sensory/perceptual effects, should affect the estimated scale values r_i, only their reliability. This is the source of the assertion that direct scaling results should be independent of the ability to resolve the stimuli scaled (cf. S. S. Stevens, 1975).

The ideal of psychophysical scaling is seldom, if ever, achieved. Judgment context has been repeatedly shown to affect psychophysical scales derived therefrom (see, e.g., Poulton, 1989). One powerful influence is the range of stimulus intensities judged. Variations in stimulus range both across continua (Teghtsoonian, 1971) and within a single continuum (see, e.g., Teghtsoonian, 1973; Teghtsoonian & Teghtsoonian, 1978) affect the exponent of the power function relating response magnitude to stimulus intensity in magnitude estimation experiments. The consensus seems to be the larger the range, the smaller the exponent—although Braida and Durlach (1972) found the exponent to vary directly with the range. Algom and Marks (1990) argued that because of their consistency with similar effects of stimulus range on matching stimulus levels, the effects of stimulus range on sensory scales must arise from the effects of context on the representation of the stimuli in the sensory system rather than (more conveniently) from response biases. One possible effect of increasing stimulus range on sensory representations is to increase their variability, as assumed by several of the theories of judgment mentioned above (e.g., Braida & Durlach, 1972, 1988; Gravetter & Lockhead, 1973). However, in all of the experiments we know of that have been performed to study the effects of range on exponents, including those just cited, stimulus range has been varied by varying the spacing between stimuli. Thus, effects of range and of spacing have been confounded. Moreover, since few of the same stimuli appeared in both large- and small-range ensembles, the precise effects of the range manipulation on responses to particular stimuli could be obtained for only a few stimulus intensities.

Stimulus range also dramatically affects intensity resolution. Gravetter and Lockhead (1973) and Nosofsky (1983) presented data which showed directly that intensity resolution estimated from absolute identification (or one-interval forced-choice) judgments for any given pair of stimulus intensities decreases as stimulus range increases. In Gravetter and Lockhead's (1973) experiments, stimulus range was varied in such a way that the physical intensities of several stimuli remained the same in narrow- and wide-range ensembles, thus deconfounding stimulus range and stimulus spacing. Nonetheless, changing intensities elsewhere in the stimulus ensemble in such a way as to increase overall stimulus range decreased resolution of the unchanged stimulus intensities. Nosofsky (1983), using another variation of a Thurstonian model, replicated Gravetter and Lockhead's (1973) results for single stimulus presentations. He also found that repeated stimulus presentations reduced stimulus noise but not judgment noise, and that both kinds of noise appeared to increase with increases in stimulus range. Similar results also were obtained by Hartman (1954) for auditory frequency resolution.

Although these clever experiments seem directly to implicate memory and judgment factors (and in Nosofsky's

[1983] case sensory factors) in the increase of response variability with stimulus range in a manner that is consistent with several theories, they do not directly demonstrate the applicability of these arguments to psychophysical scaling judgments. Prevailing scaling metatheory (e.g., Bolanowski & Gescheider, 1991) disqualifies absolute identification (one-interval forced-choice) as a scaling method, most strongly because of the identification function imposed by the experimenter and to a lesser extent because of the forced choice required among a small set of responses. A valid scaling method is supposed to allow an observer an unlimited response set, at least within a practical response range, in order to avoid biasing scale values (cf. Braida & Durlach, 1972; S. S. Stevens, 1975). More importantly, a valid scaling method is not supposed to specify the appropriate mapping from stimuli to responses, since ascertaining such a mapping is the primary object of doing the scaling experiment. Magnitude estimation (with unlimited numerical responses) and cross-modality matching or magnitude production, with a very large number of intensities on another sensory continuum as the response set, qualify as accepted methods, but absolute identification does not. Category judgment without an identification function, although it is a special case since it has a limited response set, also has its champions (e.g., Anderson, 1982; Krueger, 1989). Therefore, in spite of the arguments mentioned above as applied to absolute identification, an unequivocal demonstration of the relationship between intensity resolution and the results of direct scaling methods (i.e., power function exponents) has not been made using an accepted method under the rigorous conditions imposed by Gravetter and Lockhead (1973) or Nosofsky (1983).

THE PRESENT STUDY

The present experiments represent an attempt to provide such a demonstration and to test the prediction of the theory of fuzzy judgment mentioned above (Ward, 1979), as an example of the "categorical judgment theories." The experiments also are amenable to explanation within the framework of Schneider and Parker's (1994; Parker & Schneider, 1994) nonlinear amplifier approach (discussed further in the Discussion section). Three experiments were conducted, and stimulus range was manipulated in all three by using the method introduced by Gravetter and Lockhead (1973). Narrow-range and wide-range stimulus ensembles were created in which the middle six intensities were identical but the two least and the two most intense stimuli differed between the ensembles. Thus, comparisons between the discriminabilities among, and the responses given to, the middle six stimuli in each set could be made that were unconfounded by any differences incidental to the manipulation of stimulus range. The three experiments involved the three most popular direct scaling techniques: category judgment (without an identification function), absolute magnitude estimation, and cross-modality matching.

Experiment 1 (category judgment) served two purposes. First, category judgment is the most similar direct scaling method to absolute identification, which was used by Gravetter and Lockhead (1973). Second, as mentioned above, several authors have argued that category judgment is actually the most valid and useful of the direct scaling methods (e.g., Anderson, 1982; Krueger, 1989). If simply omitting the identification function destroys the relationship between intensity resolution and the scale values estimated from responses, then even category judgment would pass a rigorous test of scaling validity. Such a result could arise if a decrease in intensity resolution were accompanied by an increase in response variability but the mean responses to identical stimuli remained unchanged. The stimuli would then appear to be closer together in units of response variability, but they would be the same distance apart on the basis of mean responses, apart from regression effects caused by increased sampling error, which could be detected by poorer fits of the mean responses to power functions and corrected for by increasing the number of responses required.

Absolute magnitude estimation (Experiment 2) is a version of magnitude estimation in which the observer is asked to match number magnitude to sensation magnitude without any restrictions whatsoever. The observer is further admonished to make each match independent of any others made in the same series. Zwislocki and Goodman (1980) and others (e.g., Gescheider, Bolanowski, & Verrillo, 1992) have argued that this scaling method is relatively free of most biases and yields an absolute scale of sensory magnitude—that is, one where each stimulus has one and only one scale value; no scale transformations at all are permissible. Whether this contention will prove to be accepted or not, the method is a popular one and yields excellent results (e.g., Gescheider et al., 1992; Zwislocki & Goodman, 1980). If the contention is correct, changes in intensity resolution of the stimuli being scaled should be unrelated to the "true" scale values estimated from observers' responses, except for regression effects arising from increased response variability.

Cross-modality matching (Experiment 3) has been argued to be the method least likely to be affected by the ubiquitous biases attributable to humans' inconsistent use of numbers as responses (e.g., S. S. Stevens, 1975). It has also been argued to be the most natural (Zwislocki, 1991) and the most useful (Luce, 1991; Marks, 1991; J. C. Stevens, 1991; Ward, 1991) method available. It is possible that, once freed of the constraints and biases of number usage, observers can match intensities on two sensory continua in such a way that the average matches will be unrelated to changes in intensity resolution on the separate continua. Under this hypothesis, only the variability of the matches would vary with intensity resolution (again, apart from regression effects).

METHOD

Subjects

Eighteen undergraduate students at the University of British Columbia were paid to participate, 6 different people in each of Experiments 1, 2, and 3. None had any known hearing defects or any difficulty in hearing the tones used as stimuli. Four males and 2 females served in Experiment 1, 3 females and 3 males in Experiment 2, and 4 females and 2 males in Experiment 3.

Stimuli and Apparatus

The stimuli to be judged for all experiments were 1-sec-duration, 1000-Hz sinusoids generated, amplified, and electronically gated (2.5-msec rise and fall times) by a custom-built digitally controlled sound generator and timed by a Hewlett-Packard Vectra ES/12 microcomputer (Intel 80286 processor), which also recorded subjects' responses. The tones were presented monaurally to the preferred ear through Koss Pro-4AAA earphones while subjects sat in a sound attenuation chamber. Responses were made on a standard computer keyboard. Two stimulus sets (each consisting of 10 different intensities) were presented to each subject. The middle six stimuli (intensities of 61–76 dB in 3-dB steps) were common to both sets. Two stimuli (separated by 3 dB) were added 3 dB from the stimuli at each end of this set to complete the narrow set (*55*, *58*, 61, 64, 67, 70, 73, 76, *79*, *82*—added stimuli in italics), and 18 dB from the stimuli at each end to complete the wide set (*40*, *43*, 61, 64, 67, 70, 73, 76, *94*, *97*—added stimuli in italics). All intensities were measured at the earphone with a custom-built artificial ear and a General Radio precision sound-level meter. In Experiment 3, the apparatus also included a green-yellow (565-nm) LED with rise/fall time of less than 100 μsec that was used to display the light intensities to be matched to pure tone intensities. The LED was embedded in a circular block of diffusing plastic so that the light stimulus appeared as a 1° disk of uniform luminance when the observer was seated 100 cm from the stimulus. Light intensities were controlled digitally by the computer by variably attenuating the voltage across the LED.

Procedure

Each subject served on 2 consecutive days (for about 1–1.5 h on each day) at approximately the same time of day. Half of the subjects in each experiment judged the narrow stimulus set on Day 1 and the wide stimulus set on Day 2, and the other half of the subjects judged the wide set on Day 1 and the narrow set on Day 2. Each day consisted first of a 50-trial practice run, and then of two 300-trial experimental runs, for a total of 600 experimental trials, with stimulus amplitudes presented in a random sequence. About 60 experimental judgments per stimulus amplitude were made on each day. Subjects were not aware that only 10 different amplitudes were presented in any one run, nor were they told that they would be judging a stimulus set of a wider or narrower range (depending on the subject/condition) on Day 2 than on Day 1.

In Experiment 1, subjects received the following category judgment instructions:

> In this experiment you will be asked to judge the loudnesses of tones. The judgment method involves assigning categories to match the intensities of the sensations of loudness you will experience. There will be 10 categories of sensory intensity into which your judgments could be placed, so you will assign the numbers from 1 (lowest intensity) to 10 (highest intensity) inclusive to describe the category into which a given stimulus falls. To give you an idea of the range of the stimuli, the faintest and the most intense sounds will be presented in the beginning of the experiment. Mentally call the faintest sound "1" and the loudest sound "10." Try to divide the interval between the two extremes into 10 equal categories. After you have made your judgment by pressing a number from "1" to "0" (for "10") on the keyboard, press the spacebar.

In Experiment 2, subjects were given the following absolute magnitude estimation instructions (similar to those recommended by Zwislocki & Goodman, 1980):

> In this experiment we would like to find out how loud various intensities of sound appear to you. For this purpose, you will hear in the earphones a series of tones, one at a time. Your task will be to assign a number to every tone in such a way that your impression of how large the number is matches your impression of how loud the sound is. You may use any positive numbers that appear appropriate to you—whole numbers or decimals (convert fractions to decimals). Do not think of physical units of measurement, such as decibels, and do not worry about running out of numbers—there will always be a smaller number than the smallest you use and a larger one than the largest you use. Do not worry about numbers you assigned to preceding sounds. After you have made your estimate, type in the number and press the spacebar.

In Experiment 3, subjects were given the following cross-modality matching instructions:

> In this experiment you will be asked to judge loudness of sounds by adjusting the brightness of a light to a level that seems to match the intensity of each sensation of loudness you experience. That is, if a stimulus gives rise to an intense sensation, you will adjust the light so that its brightness is relatively high; if a stimulus is very faint you will adjust the light so that its brightness is relatively low. You should try to make an "absolute match" of the brightness and the sensation intensity on each stimulus presentation.

In this experiment, the response measure was the luminance of the LED light (foot-lamberts). On each trial, immediately after presentation of the tone to be judged, the LED display initially showed a different randomly chosen luminance within the range available. Subjects then repeatedly pressed an up-arrow key (which progressively brightened the light), or a down-arrow key (which progressively dimmed the light) until they were satisfied that the brightness of the light matched the loudness of the pure tone that they had just heard. They then pressed the space bar on the keyboard to end that trial and initiate the next one.

RESULTS

For purposes of these experiments, intensity resolution of the adjacent pairs of the six identical stimuli in each ensemble was measured analogously to a Thurstonian distance measure. This distance measure was designed so that results from the magnitude estimation and cross-modality matching experiments could be compared with those of the category judgment experiment, and so that the relationship between power function exponents and intensity resolution predicted by various theories could be explored. Because there is no single correct response to a stimulus in any scaling paradigm, it is not possible to calculate d' values from scaling data. Instead, d'-like measures (D_{ij}) in all experiments were calculated by dividing the difference between the arithmetic means of the response distributions (mnR_i) for each of two adjacent stimuli by a measure of the response variability for the two stimuli of the pair. Response variability was measured by the mean standard deviation: the square root of the arithmetic mean of the variances for the two response distributions ($varR_i$). Thus,

$$D_{ij} = (mnR_j - mnR_i) / [(varR_j + varR_i) / 2]^{1/2}. \quad (3)$$

This distance measure was calculated for all adjacent stimulus pairs within the set of six stimuli that were identical for narrow and wide ranges. The resulting five D_{ij} values for each subject were summed to yield a single distance measure for each of the narrow and wide conditions. The summed D_{ij} values appear in Table 1 for each subject in each experiment. The average summed D_{ij} differed significantly between the narrow- and wide-range conditions in all three experiments [category judgment, $t(5) = 3.41$, $p < .02$; absolute magnitude estimation, $t(5) = 5.68$, $p < .005$; cross-modality matching, $t(5) = 6.58$, $p < .002$]. In the category judgment experiment, 5 of the 6 subjects had a substantially greater summed D_{ij} for the narrow-range condition than for the wide-range condition; the 6th subject had a very small difference in the opposite direction. In the other two experiments, all subjects had greater summed D_{ij} for the narrow- than for the wide-range condition.

Table 2 shows estimates of the exponents of power functions fitted to individual subjects' data in all three experiments by estimating β in the following linear regression equation applied only to the geometric mean responses (gmnR) to the six stimuli that were identical in both narrow- and wide-range conditions:

$$\log(\text{gmnR}) = \beta \log(P) + \gamma + \varepsilon, \tag{4}$$

where P is stimulus sound pressure amplitude (in dyne/cm^2), β and γ are the coefficients to be estimated, and ε is error. The regression coefficient β in Equation 4 is an estimate of the best-fitting exponent (m) of a power function of the form of Equation 5:

$$\text{gmnR} = a\,P^m. \tag{5}$$

These estimates will be referred to as "exponents" and symbolized as m in what follows, as is standard in the scaling literature. The average exponents differed significantly for the narrow- and wide-range conditions in all three experiments [category judgment, $t(5) = 3.32, p < .025$; absolute magnitude estimation, $t(5) = 5.46, p < .005$; cross-modality matching, $t(5) = 3.37, p < .02$]. All 18 subjects had smaller exponents for the wide-range condition than for the narrow-range condition.

Table 3 shows the mean adjusted r^2, which measures the fit of Equation 4 to the mean responses, for individuals in each of the experiments. The values of r^2 shown in Table 3 were adjusted to the level expected when using this

Table 1
Distance Measures Summed Over Five Adjacent Pairs for Middle Six Stimuli

Subject	Narrow	Wide	Narrow − Wide
	Category Judgment		
1	2.14	1.85	0.29
2	1.59	0.96	0.63
3	2.60	1.32	1.28
4	2.04	1.03	1.01
5	1.17	1.21	−0.04
6	2.23	1.32	0.91
M	1.96	1.28	0.68
	Magnitude Estimation		
1	3.54	2.34	1.20
2	2.35	0.94	1.41
3	3.04	1.58	1.46
4	4.09	1.39	2.70
5	2.38	1.16	1.22
6	1.69	0.83	0.86
M	2.85	1.37	1.48
	Cross-Modality Matching		
1	1.62	0.89	0.73
2	2.62	1.93	0.69
3	0.90	0.32	0.58
4	1.31	0.30	1.01
5	1.16	0.68	0.48
6	2.19	0.93	1.26
M	1.63	0.84	0.79

Table 2
Regression Coefficients From Middle Six Stimuli

Subject	Narrow	Wide	Narrow − Wide
	Category Judgment		
1	0.613	0.525	0.088
2	0.253	0.160	0.093
3	0.461	0.159	0.302
4	0.351	0.223	0.128
5	0.267	0.234	0.033
6	0.611	0.306	0.305
M	0.426	0.270	0.158
	Magnitude Estimation		
1	0.286	0.216	0.070
2	0.777	0.522	0.255
3	0.457	0.350	0.107
4	0.333	0.180	0.153
5	0.722	0.495	0.227
6	0.278	0.143	0.135
M	0.476	0.318	0.158
	Cross-Modality Matching		
1	0.885	0.659	0.226
2	1.716	1.643	0.073
3	1.376	0.683	0.693
4	0.970	0.500	0.470
5	0.552	0.349	0.203
6	0.929	0.721	0.208
M	1.071	0.759	0.312

equation in a new sample from the same population, using the following formula:

$$r^2_{\text{adj}} = r^2 - (p-1)(1-r^2)\,/\,(n-p), \tag{6}$$

where n is the number of cases (here 6) and p is the number of predictors (here 2).

Figures 1, 2, and 3 display the arithmetic means of subjects' geometric mean responses to the various sound levels (i.e., average psychophysical functions) for the narrow- and wide-range conditions for category judgment, absolute magnitude estimation, and cross-modality matching, respectively. Plots for individual subjects are not shown. Data of individual subjects were highly similar to each other and to the plots shown here but departed somewhat more from the best-fit line in most cases. The lines through the data points in Figures 1, 2, and 3 are the best-fitting (using linear regression) functions of the form of Equation 4 with parameters estimated only from the responses to the middle six stimuli in each set. The estimated m for this function for the narrow-range condition was greater than that for the wide-range condition in all three experiments: category judgment, 0.411 and 0.224 for narrow- and wide-range conditions, respectively; magnitude estimation, 0.483 and 0.324; cross-modality matching, 1.017 and 0.759. The (adjusted) r^2 for these group functions are shown in Table 3; they are very close for the two conditions for all experiments.

It is possible that the differences in estimated power function exponents in narrow- and wide-range conditions in the three experiments were caused by statistical regression. This could happen if the response variability were greater in the wide- than in the narrow-range condition. Since the

Table 3
Adjusted r^2 Measures of Goodness of Fit

	Mean Individual		Group Function	
Experiment	Narrow	Wide	Narrow	Wide
Category judgment	.972	.943	.995	.995
Magnitude estimation	.980	.926	.994	.975
Cross-modality matching	.931	.904	.989	.987

number of responses per stimulus was the same, estimates of the mean response to each stimulus would be less stable when response variability was larger, and the means would be expected to cluster less closely about the best-fit line. It can be seen in Figures 1, 2, and 3 and Table 3 that this is not the case: fits appear to be about as good in wide-range as in narrow-range conditions in all three experiments. Combined with the obvious changes in the average responses to the middle six stimuli in all three experiments, but particularly for magnitude estimation and cross-modality matching, these group functions and the average r^2 values for individual functions provide little evidence that the exponent difference between narrow- and wide-range conditions was caused solely by a regression effect arising from greater response variability in the wide-range condition. Thus, the exponent differences reflect real differences in scale values obtained under the two conditions. Moreover, claims that absolute magnitude estimation or cross-modality matching produce absolute scale values are not supported by these results.

Also notice in each of Figures 1, 2, and 3 that for the narrow-range points the line fitted to the reponses to the middle six stimuli also fits equally well the responses to the two most extreme stimuli at each end of the range. However, as well as being shallower, the function for the middle six stimuli of the wide-range condition does not fit responses to the extreme stimuli well. A considerably steeper function (but still less steep than that for the narrow-range stimuli) would be required in order to fit those stimuli. Moreover, in Figures 2 and 3, and unlike the results for category judgment, the overall mean responses to the middle six stimuli are smaller in the wide- than in the narrow-range condition. Thus, the difference between the conditions in the intensities of the most extreme stimuli influenced the scale values estimated for these stimuli in a way not conceivably attributable to statistical regression.

Another interesting aspect of the plot in Figure 1 is that the responses used for the extreme stimuli differ for those for the two stimulus sets, with more extreme responses used in the wide-range condition. This represents a failure of subjects to follow the category judgment instructions, which required them to use the extreme responses for the extreme stimuli regardless of their absolute magnitudes. Such a failure represents an influence of what Ward (1987) called an *absolute*$_I$ scale on subjects' category judgments. In other words, the category judgments are a compromise between relative and absolute judgment, a finding typical of scaling judgments (cf. Marks, Szczesuil, & Ohlott, 1986).

DISCUSSION

The results described in the previous section have demonstrated that in spite of prevailing direct psychophysical scaling metatheory and practice, measured scale values of subjective magnitude can be affected by overall stimulus range even when the particular stimulus intensities under consideration remain constant. Thus, the claim that such direct scaling methods give rise to invariant scale values for psychological magnitudes (except for response biases) is not supported by these results.

The present results may perhaps best take their place alongside the plethora of context effects on scaling judgments that have been studied in the past decade (see, e.g., Marks, 1988; Schneider & Parker, 1990). All of these results and more place severe constraints on the applicability of such scales, although they do not obviate them. In fact, they argue for the adoption of a set of standard scales, and a set of standard conditions for achieving them, against which such context effects and other biases of judgment can be assessed (cf. West & Ward, 1994).

The results presented here provide support for the several theories that have proposed that all intensity judgment is categorical at base (e.g., Braida & Durlach, 1988; Gravetter & Lockhead, 1973; Thurstone, 1927; Ward, 1979). The effects of stimulus range on intensity resolution in the present experiments were highly similar to those observed for the absolute identification (or one-interval forced-choice) paradigm (but cf. Braida & Durlach, 1972). Such effects are not predicted by any theory of scaling judgments that depends only on a mapping from internal stimulus representations to responses, unless the stimulus representations are assumed to be affected by all of the stimuli in the

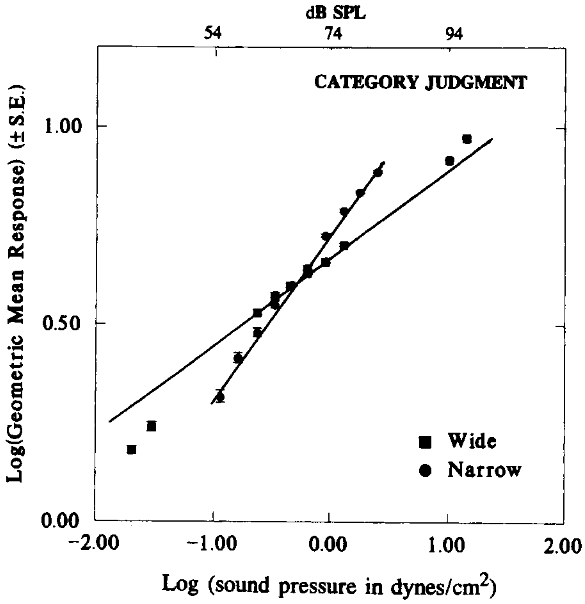

Figure 1. Psychophysical functions from Experiment 1: category judgment. The best-fitting lines plotted are based only on the middle six stimuli in each set.

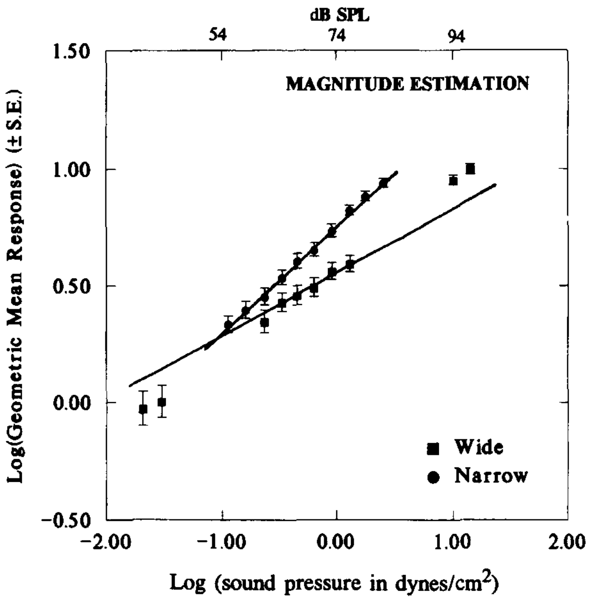

Figure 2. Psychophysical functions from Experiment 2: absolute magnitude estimation. The best-fitting lines plotted are based only on the middle six stimuli in each set.

ensemble to be judged (not a usual assumption, but see Algom & Marks, 1990). Finding such effects with an avowedly categorical method like category judgment, or with a method subject to number biases like absolute magnitude estimation, is not really surprising. However, finding the effects on scaling judgments in cross-modality matching, in whose paradigm there is not the slightest hint of a requirement to categorize intensities, is significant. Clearly, both the way we think about such "absolute matching" (cf. S. S. Stevens, 1975) and the way we relate our empirical results to measurement theory (cf. Ward, 1990) will be affected if the "best" an observer can do is to match categories.

Although it is tempting to take the present results as supporting theories of intensity judgment based on categorization, that is not the only possible explanation of these data. For example, Schneider and Parker (1994; Parker & Schneider, 1994) have proposed a nonlinear amplifier under top-down control as an early stage in auditory processing. In order to optimize intensity resolution, the gain of the amplifier is set so as to avoid peak-clipping distortion at the upper end of the range of stimulus intensities to be resolved, and so is a function of the maximum stimulus intensity encountered in a given situation. In this theory, the gain of the amplifier affects the sensory intensity that is input into both intensity resolution and loudness judgment processes and thus affects both in a correlated way. Whenever both intensity resolution and sensation magnitude are measured under conditions in which a single amplifier gain setting must be used for all stimuli judged, worse intensity resolution should be accompanied by a shallower psychophysical function (lower exponent). That is the case in the present experiments, since both psychophysical functions and distance measures of intensity resolution were derived from the same judgments, and since stimulus intensity varied randomly from trial to trial, preventing subjects from setting the amplifier gain optimally for each presented stimulus intensity. Instead they would have been forced to choose a single gain setting, one that was as high as possible without distorting the most intense stimulus in the ensemble, for use on all trials. This means that they would have been able to set the gain higher for the narrow-range condition than for the wide-range condition, since the most intense stimulus is lower for the narrow-range condition. This higher setting yields better intensity resolution (and steeper psychophysical functions) for all stimuli in the narrow-range condition, including those that remained the same across conditions. Clearly this prediction is confirmed in the present data.

Another prediction of the nonlinear amplifier theory is that the ability to resolve adjacent pairs of stimuli will improve with increasing stimulus amplitude (Schneider & Parker, 1994). Figure 4 shows the average (across 6 observers) D_{ij} values for the five adjacent pairs of stimuli for each condition of all three experiments. It is clear that D_{ij} values in the wide-range conditions tend to remain constant with stimulus intensity, whereas there is a noisy trend for them to increase with stimulus intensity for the narrow-range conditions (although none of the linear regressions yielded significant slopes). These stimuli are different enough so that an approximately 2-dB difference in differential threshold should exist between the extreme pairs in both conditions. It would seem that something more must be going on than just nonlinear amplification, but it's not clear what that is. At any rate, an experiment that manipulates stimulus range by only downward (in intensity) extensions should be done, since the nonlinear amplifier the-

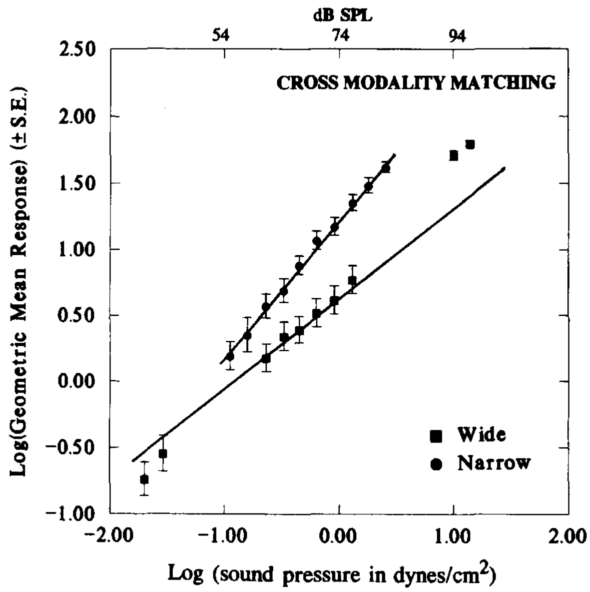

Figure 3. Psychophysical functions from Experiment 3: cross-modality matching. The best-fitting lines plotted are based only on the middle six stimuli in each set.

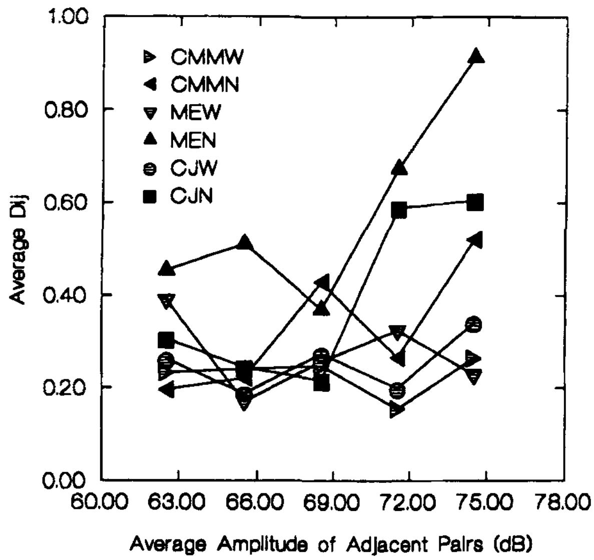

Figure 4. Mean D_{ij} values for adjacent pairs of stimuli in both conditions of all three experiments. CJ, category judgment; ME, absolute magnitude estimation; CMM, cross-modality matching brightness to loudness; N, narrow-range condition; W, wide-range condition.

ory predicts no effects of stimulus range under such a manipulation. If comparable results to the present ones were found, then the categorization explanation of the present data would gain force.

A question begged by the present results is their implications for the ongoing discussion of the relationship between the discriminability of sensory intensities (in the sense of difference threshold) and their subjective magnitude. Ever since Fechner (e.g.,1860/1966), psychophysicists have debated the nature of this relationship. The implications of this work for that debate depend on how the relationship between intensity discriminability and intensity resolution is construed. Several different relationships are possible, each with different implications. These will be outlined in what follows.

First, there are reasons to maintain a clear distinction between the two concepts. For example, Marks (1995) suggested that discriminability represents "the 'capacity' of a sensory/perceptual system to respond differentially (to) a pair of stimuli" (p. R-1). This capacity is measured in particular paradigms (e.g., two-interval forced choice) under ideal conditions that minimize the influence of "central" mechanisms such as memory or decision making, and it is intended as a descriptive property of a sensory/perceptual system. On the other hand, the extent to which particular responses are actually assigned to particular stimuli represents a more "operational" property of the entire organism and is readily influenced by memory load, decision load, and other factors that involve central mechanisms. From this perspective, the present results, which emphasize the more operational property, have no implications at all for the relationship between discriminative capacity and subjective magnitude.

The complementary position, that the two concepts are in fact identical, has also been promoted. For example, Thurstone (e.g., 1927) and Garner (e.g., 1952) both based scaling of subjective magnitudes on the fact that categorical responses to the same stimulus intensity presented on different occasions vary, and this response variability in turn varies with stimulus intensity similarly to Weber's law. Taking a measure of this response variability as equivalent to a jnd measure of discriminability and using a summation process similar to Fechnerian integration results in a scale of psychological magnitude that closely resembles those based on the jnd measures (Garner, 1952, 1958). Thurstone (1927) and Garner (1952, 1958) apparently felt that measures of the variability of responses to a fixed stimulus, or of the inability of subjects consistently to assign different responses to different stimuli, in paradigms such as absolute identification (one-interval forced choice) under less than ideal conditions (high memory load, larger stimulus range, very low or high stimulus intensities, etc.) also should be considered to be measures of discriminability. Given this position, the present results imply that assessment of subjective magnitude is profoundly influenced by the discriminability of the stimuli to be judged.

Yet a third position is possible. It is best illustrated by Braida and Durlach's (1988) theory of intensity resolution. In that theory, d' is directly proportional to intensity differences and inversely proportional to a variance term arising from both peripheral (sensory) and central (memory, judgment) sources. The value of d' observed in an experiment is maximal whenever the variance term is minimal. However, "even simple discrimination tasks require the use of central mechanisms for intensity comparisons, and it is important to factor out the limitations imposed by these mechanisms in drawing conclusions about peripheral encoding" (Braida & Durlach, 1988, p. 581). The "discriminative capacity" of a sensory system would seem to be an inference based on an extrapolation of what d' would be if variance from central sources were eliminated. In this view, all we ever have is an "operational" measurement of whole-organism performance, from which we must infer properties of the sensory system. Certain paradigms might give us direct estimates of what Marks (1995) called "discriminative capacity," but such estimates would still require some extrapolation. Given this view, both discriminative capacity and subjective magnitude would be expected to covary with intensity resolution, since both are derivative from the same judgment and sensory mechanisms, similarly to the nonlinear amplifier theory of Schneider and Parker discussed above. However, since there is no consensus about which of the views described above should prevail, the implications of the present work for the discriminability–subjective magnitude debate remain unclear.

REFERENCES

Algom, D., & Marks, L. E. (1990). Range and regression, loudness scales and loudness processing: Toward a context-bound psychophysics. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 706-727.

ANDERSON, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

BOLANOWSKI, S. J., JR., & GESCHEIDER, G. A. (EDS.) (1991). *Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens*. Hillsdale, NJ: Erlbaum.

BRAIDA, L. D., & DURLACH, N. I. (1972). Intensity perception II: Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, **51**, 483-502.

BRAIDA, L. D., & DURLACH, N. I. (1988). Peripheral and central factors in intensity perception. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 559-583). New York: Wiley.

DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception I: Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.

FECHNER, G. T. (1966). *Elements of psychophysics* (Vol. 1, H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

GARNER, W. R. (1952). An equal discriminability scale for loudness judgments. *Journal of Experimental Psychology*, **43**, 232-238.

GARNER, W. R. (1958). Advantages of the discriminability criterion for a loudness scale. *Journal of the Acoustical Society of America*, **30**, 1005-1012.

GESCHEIDER, G. A. (1985). *Psychophysics: Method and theory* (2nd ed.). Hillsdale, NJ: Erlbaum.

GESCHEIDER, G. A., BOLANOWSKI, S. J., JR., & VERRILLO, R. T. (1992). Sensory, cognitive and response factors in the judgment of sensory magnitude. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 575-621). Amsterdam: Elsevier, North-Holland.

GRAVETTER, F., & LOCKHEAD, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review*, **80**, 203-216.

HARTMAN, E. B. (1954). The influence of practice and pitch-distance between tones on the absolute identification of pitch. *American Journal of Psychology*, **67**, 1-14.

KRUEGER, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral & Brain Sciences*, **12**, 251-320.

LUCE, R. D. (1991). What is a ratio in ratio scaling? In S. J. Bolanowski, Jr. & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens* (pp. 8-17). Hillsdale, NJ: Erlbaum.

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

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

MARKS, L. E. (1995). [Unpublished review of a draft of this paper]

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

NOSOFSKY, R. M. (1983). Shifts of attention in the identification and discrimination of intensity. *Perception & Psychophysics*, **33**, 103-112.

PARKER, S., & SCHNEIDER, B. (1994). The stimulus range effect: Evidence for top-down control of sensory intensity in audition. *Perception & Psychophysics*, **56**, 1-11.

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

SCHNEIDER, B., & PARKER, S. (1990). Does stimulus context affect loudness or only loudness judgments? *Perception & Psychophysics*, **48**, 409-418.

SCHNEIDER, B., & PARKER, S. (1994). Stimulus context effects and the psychophysical law: Is coexistence possible? In L. M. Ward (Ed.), *Fechner Day 94* (pp. 71-76). Vancouver, BC: International Society for Psychophysics.

STEVENS, J. C. (1991). Magnitude matching: Application to special populations. In S. J. Bolanowski, Jr. & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens* (pp. 43-58). Hillsdale, NJ: Erlbaum.

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

TEGHTSOONIAN, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71-80.

TEGHTSOONIAN, R. (1973). Range effects in psychophysical scaling and a revision in Stevens' law. *American Journal of Psychology*, **86**, 3-27.

TEGHTSOONIAN, R., & TEGHTSOONIAN, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, **24**, 305-314.

THURSTONE, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 273-286.

WARD, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 444-459.

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

WARD, L. M. (1990). Cross-modal additive conjoint structures and psychophysical scale convergence. *Journal of Experimental Psychology: General*, **119**, 161-175.

WARD, L. M. (1991). Associative measurement of psychological magnitude. In S. J. Bolanowski, Jr. & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens* (pp. 79-100). Hillsdale, NJ: Erlbaum.

WEST, R. L., & WARD, L. M. (1994). Constrained scaling. In L. M. Ward (Ed.), *Fechner Day 94* (pp. 225-230). Vancouver, BC: International Society for Psychophysics.

ZADEH, L. A. (1965). Fuzzy sets. *Information & Control*, **8**, 338-353.

ZWISLOCKI, J. J. (1991). Natural measurement. In S. J. Bolanowski, Jr., & G. A. Gescheider (Eds.), *Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens* (pp. 18-26). Hillsdale, NJ: Erlbaum.

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

(Manuscript received March 3, 1995;
revision accepted for publication October 4, 1995.)