

# On theoretical and realizable ideal conditions in psychophysics: Magnitude and category scales and their relation\*

HANNES EISLER

*Psychological Laboratories, University of Stockholm, Box 6 706, S-1 13 85 Stockholm, Sweden*

and

HENRY MONTGOMERY

*University of Göteborg, Göteborg, Sweden*

By spacing 10 stimuli (white noise) between 40 and 110 dB according to two criteria [equal response ambiguity (ERA) and equal discriminability (ED)], an attempt was made to construct an "ideal" case for magnitude estimation and category rating. The "ideal" case is defined by linear and constant Weber functions (SDs as a function of scale values) for the two scales, respectively. Altogether, three group and two individual magnitude and category rating experiments were run with these two spacings. It was found that the ERA spacing approximated the ideal case well for both Weber functions and the ED spacing only for the Weber function of the category scale. The general psychophysical differential equation that relates scale values and Weber functions for the two scales allowed good prediction of the category scales from the magnitude scales and the Weber functions. The data suggested a distinction between phenotypic (empirical) and genotypic Weber functions, analogous to "real" and "ideal" cases in physics.

Fechnerian integration is a scaling method by which from a given scale with known Weber function (uncertainty as function of scale value) another scale with constant Weber function is constructed. The process might be seen as a scale transformation (cf. Luce & Galanter, 1963). Originally, the starting scale was some physical scale (say, weight in grams) and its Weber function, Weber's law, typically obtained by the method of constant stimuli. However, being only a scale transformation method, Fechnerian integration is not restricted to physical scales as the starting point; one may well construct a new scale (with constant Weber function) by transforming a subjective scale like the magnitude scale for loudness. Now the interesting question can be posed whether this new scale agrees with one obtained by an independent procedure empirically. As shown by Eisler, theoretically (1962b) and empirically for loudness of white noise (1962a), smell (1963b), and length of line (1963c), these new scales were very close to the category scales (mean ratings) for the corresponding continua. In all these cases, the Weber functions were obtained from intraindividual SDs of the magnitude estimates, implying a generalization of the concept of Weber function to subjective continua, and furthermore to the SD as the measure of uncertainty.

In Eisler (1963a), the concept of Fechnerian integration was extended to the "general psychophysical differential equation" (GPDE)<sup>1</sup>

$$\frac{dy}{dx} = \frac{\sigma_y(y)}{\sigma_x(x)}, \quad (1)$$

\*This work was supported by the Swedish Council for Social Science Research. Free computer time was made available by the Swedish Office for Rationalization and Economy. We are indebted to Björn Koch, Maria Backlund, and Ivar Paulson for experimental and computational assistance.

where  $x$  and  $y$  denote scale values of subjective variables (central tendencies of estimates) for the same set of stimuli, and  $\sigma_x(x)$  and  $\sigma_y(y)$  the Weber functions computed as (intraindividual) SDs for the estimates.

Unlike Fechnerian integration, the GPDE is not just a scale transformation method that an E may apply; rather, it expresses a relation between subjective scales and their Weber functions for the same set of stimuli. The transformation  $y = f(x)$  is obtained by integrating Eq. 1.

$$\int_{y_0}^y \frac{dy}{\sigma_y(y)} = \int_{x_0}^x \frac{dx}{\sigma_x(x)}. \quad (1a)$$

(The limits  $x$  and  $y$  correspond to the same stimulus; subscript 0, to the lowest one.) Note the implicit invariance requirement of Eq. 1. When different procedures are used for the scaling of one of the variables, say  $x$ , e.g., by using different standards in magnitude estimation, both the scale values and the Weber functions are affected. Thus, different  $x$  scales will be obtained, but  $\sigma_x(x)$  will be different, too, leaving the right member of Eq. 1a invariant and equal to the unchanged left member.

The validity of the GPDE can be tested empirically: Scale  $y$  can be computed from Scale  $x$  and the Weber functions and compared to the empirically obtained Scale  $y$ . In this way, Eq. 1 was confirmed for a number of (empirically obtained) combinations of Weber functions, viz, constant-constant, linear-linear, linear-parabolic, parabolic-parabolic (Eisler, 1965a).

Still, the case with a constant Weber function is of particular interest, since it implies that equal probabilities of confusing two stimuli correspond to equal subjective distances, provided certain other

assumptions of the scaling model are valid (cf. Torgerson, 1958, Chap. 10). Therefore, a scale with constant Weber function can be regarded as a discrimination, equiprecision (Helm, Messick, & Tucker, 1961) or equal discriminability (Garner & Hake, 1951) scale (cf. also Eisler, 1965b).

In most of the above-mentioned studies regarding the relation between magnitude and category scales, the Weber function of the magnitude scale proved linear (except that the two SDs corresponding to the two extreme stimuli were too small) and the prediction of the category scale under the assumption of its Weber function being constant, i.e., by Fechnerian integration, came very close. This suggested two conclusions: (1) the GPDE gives a valid description of the relation between the two scales, and (2) the category scale is a discrimination scale. However, there is an empirical flaw in the argument: the Weber function for the category scale is constant only as a very crude approximation. The SDs of the ratings are typically greatest in the middle of the stimulus range and decrease towards both ends (see Fig. 12A).

Instead of rejecting the Fechnerian model because of this discrepancy, the Weber function of the category scale may be regarded as subject to distortion or bias. The distortion probably has several causes, e.g., the curtailment of possible responses to the extreme stimuli. Another possible source of perhaps greater interest is that the more extreme stimuli are recognized more often than stimuli which are closer to the middle of the stimulus range, and thus are assigned more often to the same category; the same process is likely to operate for the two extreme stimuli in magnitude estimation. This "end effect" thus seems to depend on the experimental procedure (cf. also Siegel, 1972; Siegel & Siegel, 1972). In an *ideal case* (which is impracticable), the range of the stimuli should be infinite, as should their number. An analogy from physics is provided by the magnetic field due to a solenoid. The simple formula for magnetic fields holds only for solenoids infinitely long and thin. The magnetic field of a circular coil of finite size cannot be described accurately by a simple formula, and even for very long and thin real coils there are deviations, particularly at the ends. The concept of the ideal case is a matter of course for physicists, whereas few psychologists seem to be familiar with the distinction between ideal and real cases.

The aim of the present series of experiments was to subject the assumption of a biased Weber function for the category scale to a closer study by investigating the magnitude and category scales, their Weber functions, and the relation between them for an approximation of the ideal case. This case was to be achieved by manipulating the stimuli, the greater recognizability of the outer stimuli being counteracted by moving them closer together to decrease their discriminability. Hopefully, this would eliminate or at least weaken the end effects. Two procedures for obtaining the "ideal

case" were attempted: (a) constructing a set of stimuli that are equally discriminable in accordance with Garner and Hake (1951), the ED spacing, and (b) constructing a set of stimuli with equal response ambiguity (see Attneave, 1959) for all stimuli, the ERA spacing.

The rationale for the ERA spacing was to make the stimuli equally identifiable "by definition" and thereby remove the end effect arising from the greater recognizability of outer stimuli. Thus, this procedure aims more directly at a constant Weber function for the category scale. The ED spacing, on the other hand, attempts to achieve equal identifiability by means of equal discriminability, i.e., somewhat indirectly. However, the ED spacing has the advantage that it can be constructed by a systematic procedure, whereas the ERA spacing has to be attained by trial and error.

The aim of the present study was thus to achieve by *experimental manipulation* what was done before by assumption: an approximation of the ideal case (constant Weber function of the category scale).

The following predictions were made: (1) For one or both of the above-mentioned spacings, the Weber function of the category scale is constant. (2) For one or both spacings, the Weber function of the magnitude scale likewise constitutes the ideal case: it is linear throughout, without deviations for the end points. (3) The category scale can be calculated by means of the GPDE for both spacings.

These predictions were tested for group and individual data, and for 7- and 15-point category scales for the continuum of white noise.

## GENERAL DESIGN AND APPARATUS

Figure 1 gives the experimental design with the denotations for the experiments used in the rest of the paper.

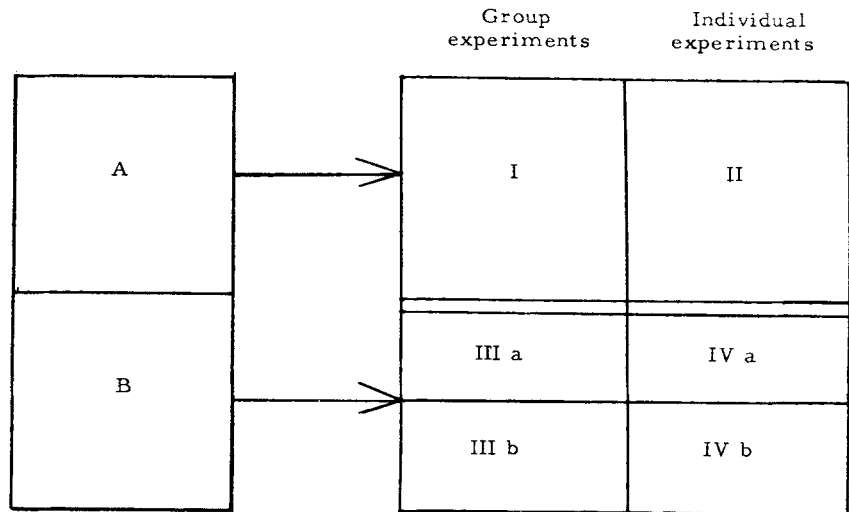
The same apparatus was used in all experiments. It consisted of a white-noise generator, adjusted to an SPL of 110 dB (re 20 microbar/m<sup>2</sup>) that fed a pair of earphones (Beyer, DT 48, 5 ohms) via a bandpass filter (75-2,400 Hz), an attenuator (Marconi, smallest step: .1 dB), a transformer, and a switch.

The two extreme stimuli were in all experiments 40 and 110 dB. In all experiments, the O presented himself with the stimulus by pressing a button after indication of E. He could listen to the stimulus as long as he wished by keeping the button pressed. After releasing the button, O gave his response orally. In Experiment A, two Os were used for each spacing. In Experiment B, the same four Os were used throughout. In the main experiments, different Os were used in each experiment (Experiments I, II, IIIa, IIIb, IVa, and IVb). Twelve Os participated in each group experiment and four Os in each of the individual main experiments.

## CONSTRUCTION OF AN IDEAL SPACING: EXPERIMENTS A AND B

At the start of each experiment, 10 noise intensities between 40 and 110 dB were presented to the O first in ascending and then in descending order. The 10 stimuli were denoted by the numbers 1-10 in ascending order. After the presentation of each stimulus, the O was told its number, the number by which he had to identify each stimulus in the experiment proper. There, the stimuli were presented 1,000 times in 100 blocks of 10

Fig. 1. Experimental design. The letters denote preparatory experiments leading to ERA spacing (A) and ED spacing (B). The main experiments are denoted by Romans. Each experiment consisted of two subexperiments, viz, one magnitude estimation experiment and one category rating experiment. In Experiments I, II, IIIa, and IVa, a 7-point category scale was used, and in Experiments IIIb and IVb, a 15-point category scale was used.



presentations. Each intensity was thus judged 100 times. Within each block, the order of the 10 stimuli was randomized with different orders for different blocks. In Experiment A, the same total order of the 1,000 presentations was used for half of the Os and the reversed order for the rest. In Experiment B, different orders were used for each O. The experiment was divided into two sessions (2-3 h/session) for each O, with a few days between the sessions. In each session, there were a few 5-min rest breaks.

**Experiment A (ERA Spacing)**

The first spacing was chosen so that the distance between adjacent stimuli was the same as on an ED scale reported by Garner and Hake (1951). The response ambiguities,  $H_x$ , were calculated according to  $H_x = -\sum p_i \log_2 p_i$ , where  $x$  denotes a stimulus,  $i$  a response category with the relative frequency  $p_i$ , and the summation is carried out over all response categories used for Stimulus  $x$ . Plots of the response ambiguities for each of the two Os separately were used as a guideline for the next spacing. This procedure was repeated with two new Os for each spacing four times, i.e., for altogether five spacings. The decibel values of the stimuli for the five spacings are given in Table 1. The

fourth spacing (Fig. 2) agreed most with the criterion, constant response ambiguities. For both Os, the response ambiguities were somewhat lower for the outer stimuli than for the rest, more markedly so for the softer (left) end. An attempt to counteract this drop at the ends by crowding the outer stimuli still more (Spacing 5) failed; as Fig. 2 shows, here the *middle* stimuli had the lowest response ambiguities, entailing an increase in the deviation from constancy. It is noteworthy that this slight change in spacing (cf. Spacing 4 and Spacing 5 in Table 1) causes such a drastic change in the relative sizes of the response ambiguities.

Thus, Spacing 4 was used in the main experiments.

**Experiment B (ED Spacing)**

The first spacing was chosen by means of ED scales that were computed from three of the response frequency matrices obtained in Experiment A (Spacings 3, 4, and 5) and from a response frequency matrix obtained from a spacing with equal intervals in decibels between successive stimuli. The means of the decibel

Table 1  
Spacings in Decibels Used in Experiments A and B

	Stimulus Number										SD*		
	1	2	3	4	5	6	7	8	9	10			
Experiment A:													
Spacing 1	40	45	50	57	69	81	93	100	105	110	.47	.21	
Spacing 2	40	41	43	47	55	80	96	104	108	110	.63	.23	
Spacing 3	40	41	43	47	55	95	103	107	109	110	.20	.23	
Spacing 4	40	42	47	52	62	88	98	103	108	110	.17	.20	
Spacing 5	40	41	43	51	66	84	99	107	109	110	.28	.31	
Experiment B:													
First Spacing	40	45	53	64	76	87	95	101	106	110	92	287	
Best Predicted Spacing for O L.S.	40	44	54	67	80	92	100	105	108.5	110	16		
Best Predicted Spacing for O T.A.	40	45	54	65	76	89	101	107.4	109.4	110	11		
Final Spacing	40	45	54	66	78	91	101	106	109	110			

Note—Spacing 4 in Experiment A was the ERA spacing used in Experiments I and II and the final spacing in Experiment B was the ED spacing used in Experiments III and IV.

\*SDs of the 10 response ambiguities for each of the two Os for the different stimulus spacings.

†Sum of squared deviations between the spacing used and the successive predicted spacing for the two Os who had random order presentation of the stimuli (Os L.S. and T.A.).

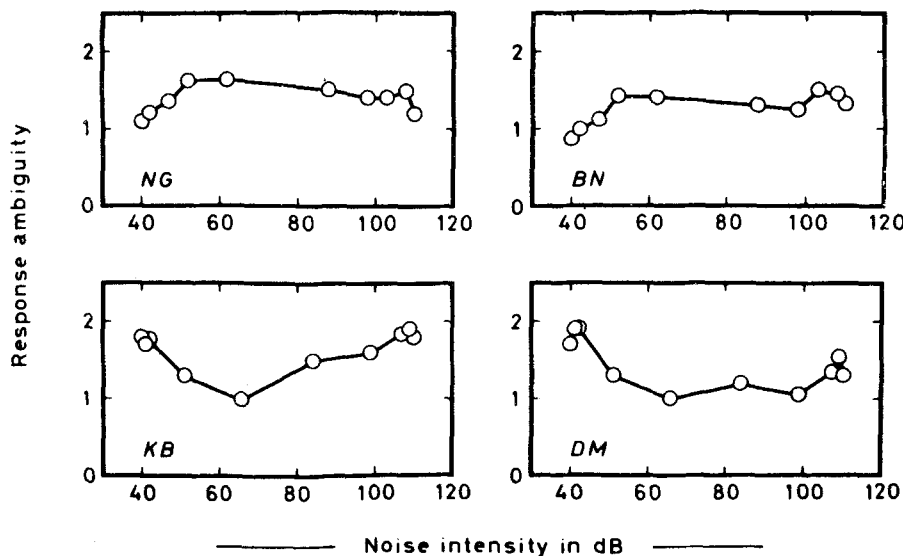


Fig. 2. Response ambiguities for two spacings of noise intensities (Spacing 4, Os N.G. and B.N., and Spacing 5, Os K.B. and D.M.). The spacings are given in Table 1. Spacing 4 was used in Experiments I and II.

values corresponding to equal intervals in these four ED scales constituted the first spacing. Four Os participated in the experiment, their ED scales were computed, the spacing changed in accordance with these results, and the procedure was repeated, three times for three of the Os and four times for the fourth O. For two of the Os, the stimuli were not presented randomly in blocks of 10. Instead, they were presented in ascending or descending series with 5, 6, or 7 stimuli in each series. We thought that this method of presentation would prevent or at least weaken the recognition of the extreme stimuli, cf. Eisler and Ottander (1963). The order between ascending and descending series was rotated according to the abba baab scheme.

Garner and Hake's method seemed to work quite well. Already the second spacing for each O was quite similar to the best spacings obtained and there did not seem to be any systematic discrepancies between the spacings obtained. As a measure of the goodness of a spacing, we used the sum of squared deviations in decibels between the spacing in question and the successive spacing obtained. Our criterion for stopping iteration was that no further decrease in the sum of squared deviations could be seen.

The sum of squared deviations showed that the Os who had random order presentation generally yielded the best results, contrary to the expectation that presenting the stimuli in ascending and descending series would do better. Therefore, for the spacing that was to be used in the main experiments, only the two Os with random presentation were used, and the means (in decibels) of the best spacings obtained for these Os constituted the final spacing (cf. Table 1).<sup>2</sup>

#### MAGNITUDE ESTIMATIONS AND CATEGORY RATINGS

##### Method

The same apparatus and the same method of presenting the

stimuli was used as in Experiments A and B. In the magnitude estimation experiments, a stimulus of medium intensity (62 dB in Experiments I and II, 78 dB in Experiments III and IV) was presented to the O and called 10 (the standard). The O was asked to estimate a series of noise intensities so that the ratio between the numbers given and 10 reflected the ratio between the sensations of the stimulus presented and the standard. In the category rating experiments, the O was presented with the weakest and strongest intensities (40 and 110 dB in all series) and informed that they were called 1 and 7 (or 1 and 15). The Os were asked to assign to each stimulus noise an integer between 1 and 7 (or 1 and 15) so that subjective sensation intervals between successive numbers were equal. The standards were presented only once, at the beginning of the experiments. In order to check that the Os had understood the instructions, each O was asked to give magnitude estimations and category ratings of the length of a few lines drawn on a blackboard. After this, there were 10 preliminary trials with all 10 sound intensities. The stimuli were presented in random order. In the main experiment, the order of presentation of the stimuli followed the same principles as in Experiments A and B. We also tried the method of presenting the stimuli in ascending and descending series for two of the individual Os (Os E.E. and H.S.) in Experiment IV. In the group experiments, every O judged every stimulus 4 times, and in the individual experiments, 50 times. In the individual experiments, there were two or three breaks for rest of about 5 min. The same order of presentation was used for every O in Experiment I. In the other experiments, different orders were used for each O. As for Experiments I and II, half the Os performed the category rating experiment before the magnitude estimation experiment and the other Os in the reverse order. In the other experiments, all Os performed the magnitude estimation experiment before the category ratings because a scrutiny of the data from Experiments I and II made us suspect that the category rating experiments would bias the Os to restrict the range of numbers used in the magnitude estimation experiment.

##### Magnitude and Category Scales

For the group experiments (Experiments I and III), magnitude scale values were computed as medians of each O's mean for each stimulus. The medians were used because both geometric and arithmetic means seemed to deviate systematically from the medians. The category scale values were computed as arithmetic means over all Os for each stimulus. In the individual experiments,

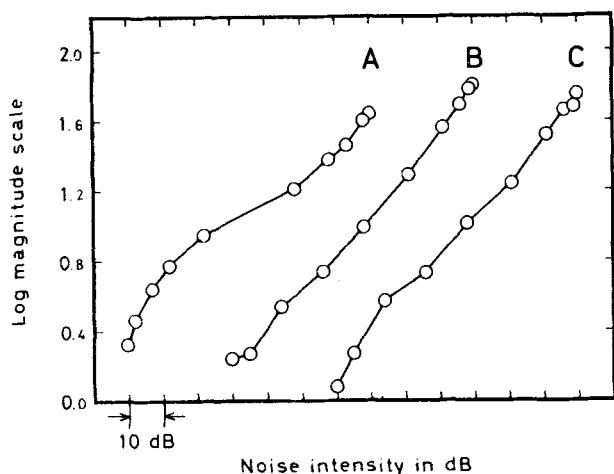


Fig. 3. Log magnitude scale for loudness as a function of noise intensity. (A) ERA spacing (Experiment I), (B) ED spacing (Experiment IIIa), (C) ED spacing (Experiment IIIb).

arithmetic means were used throughout for both scales.<sup>3</sup>

In Fig. 3, the magnitude scales from Experiments I and III are plotted in log-log coordinates against stimulus values. The scales from the two experiments are obviously differently related to stimulus values: The magnitude scale from Experiment I exhibits a reverse S form, whereas the two magnitude scales from Experiment III are approximately linear. Figure 4 shows that the category scale from Experiment I also exhibits an S form when plotted in semi-log coordinates against stimulus values. The category scales from Experiment III are both concave upward (cf. Fig. 4). The trend is more marked for the 15-point category scale. Figures 5 and 6 show that the individual scales exhibit great variations and irregularities. By and large, they do not contradict the findings for group data, though none of the individual magnitude scales follows the power law perfectly.

Figures 4 and 6 also show that the category scale intervals between successive stimuli in all experiments tend to be greatest in the middle of the stimulus range and decrease toward both ends.

Fig. 4. Category scales for loudness as a function of noise intensity. (A) ERA spacing (Experiment I), (B) ED spacing (Experiment IIIa), and (C) ED spacing (Experiment IIIb). The curves constitute predictions by the GPDE (see text).

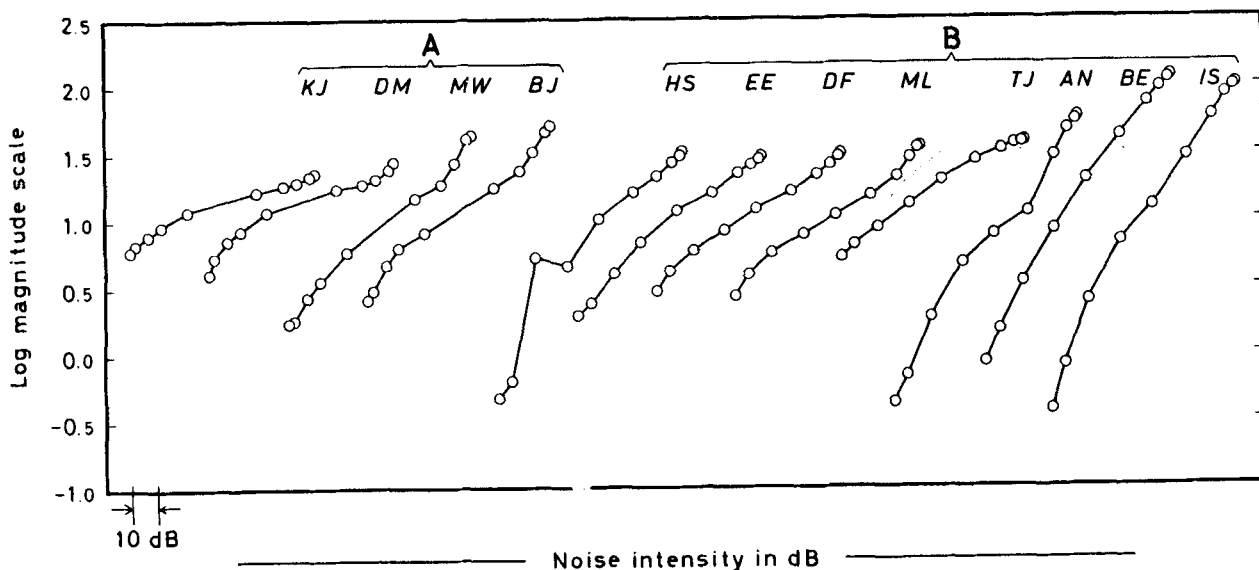
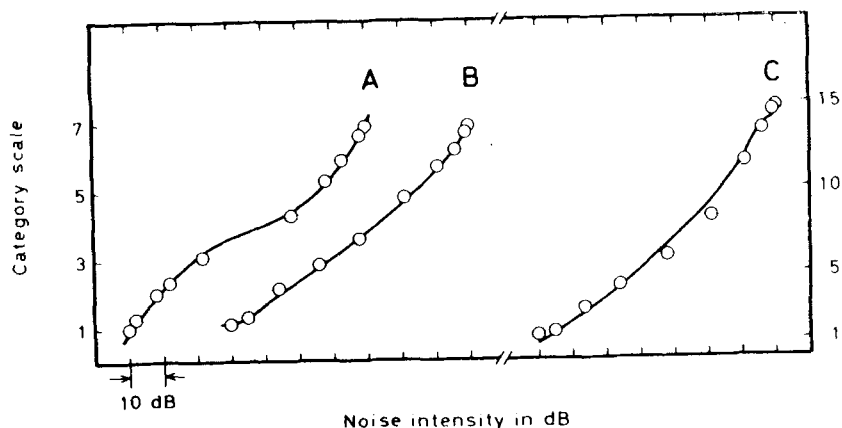


Fig. 5. Log magnitude scales for loudness for individual Os as a function of noise intensity. (A) ERA spacing (Experiment II), (B) ED spacing (Experiment IV). The Os are denoted by their initials.

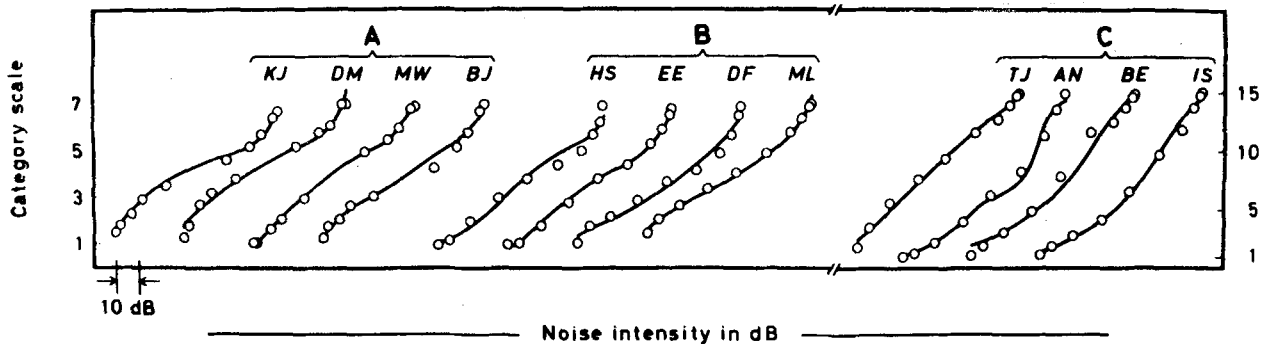


Fig. 6. Category scales for loudness for individual Os as a function of noise intensity. (A) ERA spacing (Experiment II), (B) ED spacing (Experiment IVa), (C) ED spacing (Experiment IVb). The Os are denoted by their initials. The curves constitute predictions by the GPDE (see text).

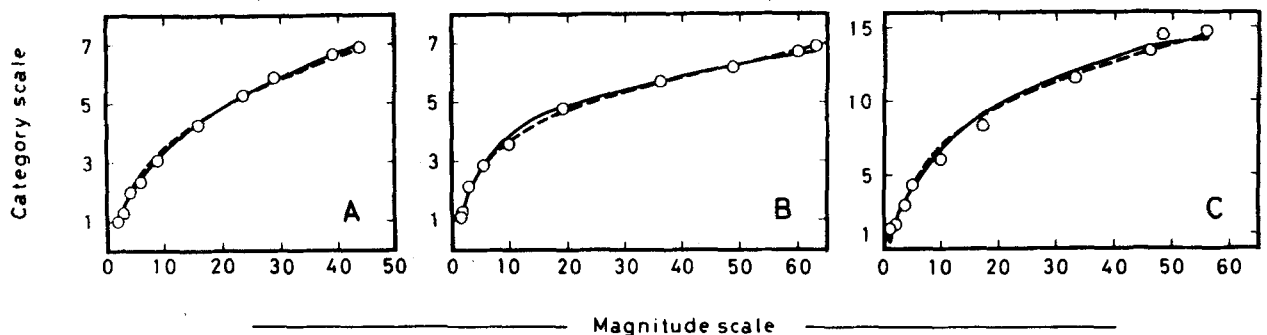


Fig. 7. Category scale values as a function of the magnitude scale for loudness. (A) ERA spacing (Experiment I), (B) ED spacing (Experiment IIIa), (C) ED spacing (Experiment IIIb). The curves constitute predictions of the category scales (solid curve: GPDE in general; dashed curve: Fechnerian integration).

The category scales obtained from the four experiments are plotted as a function of the magnitude scales in Figs. 7 and 8. The curves constitute predictions of the category scales (see section "Fechnerian Integration and the GPDE"). For group data, the function relating the category scale to the magnitude scale is concave downward. Individual data exhibit great variations over Os. The relation between the category and magnitude scale varies from almost linearity over sigmoidal relations to relations which are clearly concave downward.

In Fig. 9, the 7-point category scale from Experiment IIIa is plotted against the 15-point category scale from Experiment IIIb. The relation between the two scales seems to be weakly curvilinear.

Table 2 gives the number of judgments obtained for each category of the category scales from the group experiments. As may be seen, the extreme categories were most frequently used, indicating that the category scales were not "pure" category scale as defined by Stevens and Galanter (1957).

**Weber Functions**

The SDs used to obtain the Weber functions for group data are intraindividual SDs, computed as described by Eisler (1962a). For individual data, the SDs were

computed in the usual way. When the Weber function of the magnitude scales from the group experiments approximated linearity, the equation of the Weber function was determined by fitting a straight line to the data (cf. Fig. 10) according to a method likewise described by Eisler (1962a).

The SDs from the magnitude scales are plotted in Figs. 10 and 11. The linear case of Weber's law for the magnitude scale holds quite well in Experiment I, except that the SD corresponding to the strongest stimulus appears somewhat too small. The slope and  $\psi$  intercept on the magnitude scale (slope = 0.13 and intercept = -3.9) are close to the corresponding values obtained by Eisler (1962a) (slopes = 0.16, 0.17, and 0.19, and intercepts = -3.1, -2.7, and -3.1). The Weber functions of the two magnitude scales obtained in Experiment III deviate from the usual findings (cf. Fig. 10). The intercept of the magnitude scale is closer to zero than usual, and the function is linear for only approximately the lower half of the range. Note that this result seems to be reliable, since it was obtained with two different groups of Os.

The individual experiments exhibit various forms of Weber functions of the magnitude scale (cf. Fig. 11). In Experiment II, the Weber function is quite similar to the Weber function obtained in Experiment I for M.W. and

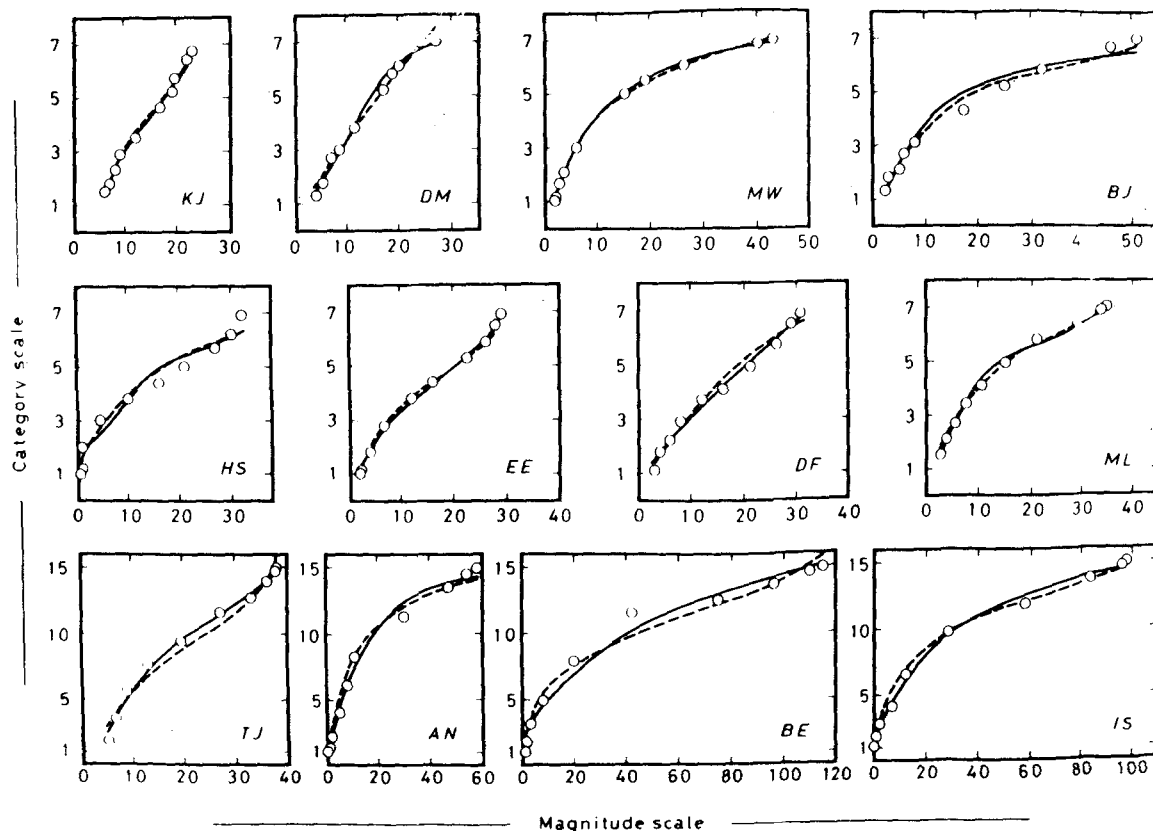


Fig. 8. Category scale values as a function of the magnitude scale for loudness for individual Os. First row: ERA spacing (Experiment I); second row: ED spacing (Experiment IVa); third row: ED spacing (Experiment IVb). The curves constitute predictions of the category scales (solid curve: GPDE in general; dashed curve: Fechnerian integration). The Os are denoted by their initials.

B.J. For K.J. and D.M., the corresponding Weber functions are almost constant. In Experiment IV, H.S.,

D.F., and A.N. exhibit approximately linear Weber functions. For the other Os in Experiment IV, the Weber functions appear more or less parabolic.<sup>4</sup> It is worth noting that the two extreme points do not deviate systematically from a linear trend for those Os who exhibited approximately linear Weber functions.

Figures 12 and 13 show plots of the SDs from the category scales. The Weber functions of the 7-point category scales for the group experiments (cf. Fig. 12) are approximately constant except for the extreme points. (Compare Panels B and C with Panel A.) The deviations from constancy are minor compared to Eisler's (1962a) results, where only the middle section of the Weber function of the category scale was approximately constant. In the individual 7-point category rating experiments, four Os exhibited approximately constant Weber functions with some exceptions for either or both extreme points (K.J. in Experiment II and M.L., D.F., and H.S. in Experiment IVa). The other Os exhibit decreasing SDs toward one or both ends. The Weber functions of the 15-point category scales for individual data (Experiment IVb) deviate conspicuously from constancy (cf. Fig. 13). The functions appear almost parabolic. The same trend, though less conspicuous, is found for group data (Experiment IIIb, Fig. 12).

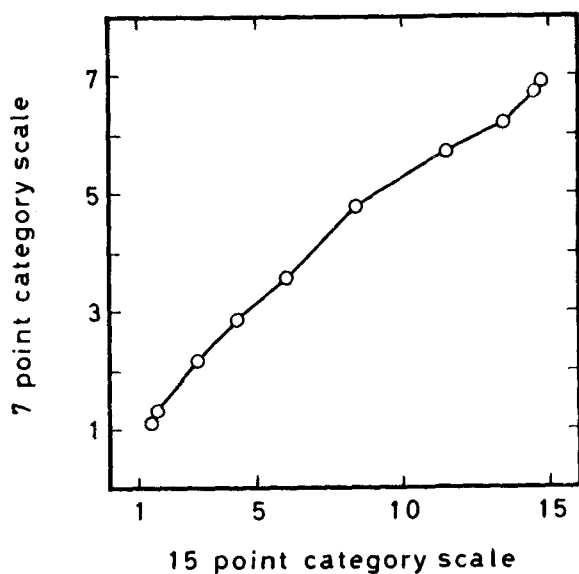


Fig. 9. The 7-point category scale from Experiment IIIa as a function of the 15-point category scale from Experiment IIIb.

**Table 2**  
**Number of Assignments for Each Category for the Category Scales from the Group Experiments (Experiments I, IIIa, and IIIb)**

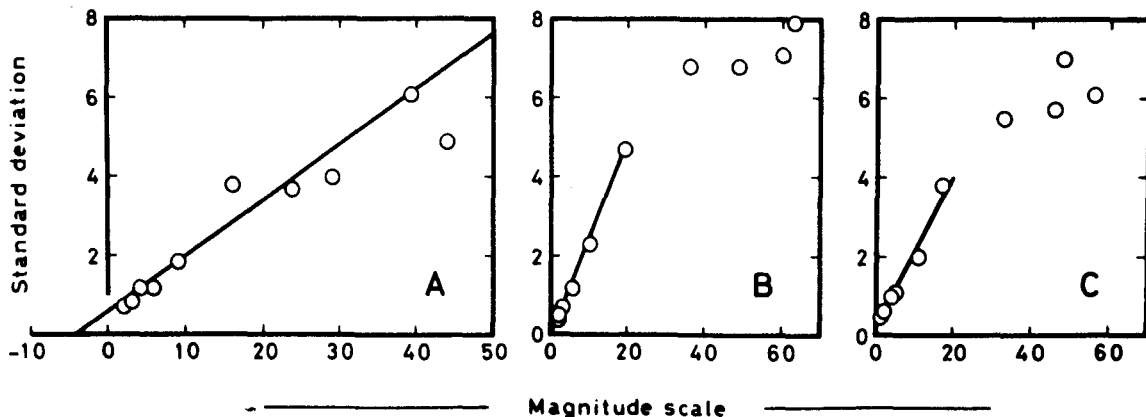
7-Point Category Scale			15-Point Category Scale	
Category	Exp I	Exp IIIa	Category	Exp IIIb
1	88	81	1	61
2	88	67	2	45
3	57	63	3	39
4	42	39	4	21
5	55	53	5	45
6	67	82	6	21
7	83	95	7	13
			8	19
			9	10
			10	27
			11	5
			12	30
			13	23
			14	52
			15	69

**FECHNERIAN INTEGRATION AND THE GPDE**

As pointed out in the introduction, Fechnerian integration can be regarded as a special case of the GPDE, viz, the case with one of the two Weber functions constant. This special case is of particular interest, since, again in agreement with arguments proposed before, the conception of the category scale as a discrimination scale requires its Weber function, at least in the ideal model, to be constant. The following data treatment allows a comparison between ideal and real Weber functions as well as a test of the GPDE in general.

For all experiments, the trapezoid method was used to calculate—point by point, with the empirical values for the 10 experimental points—the following two integrals, both derived from Eq. 1:

$$I_1 = \int \frac{\sigma_K(K)}{\sigma_\psi(\psi)} d\psi \text{ and } I_2 = \int \frac{d\psi}{\sigma_\psi(\psi)}, \quad (2)$$



**Fig. 10.** Intraindividual SDs as a function of magnitude scales for loudness. (A) ERA spacing (Experiment I), (B) ED spacing (Experiment IIIa), (C) ED spacing (Experiment IIIb). The straight line in Fig. 10A is fitted according to a method described by Eisler (1962a). The straight lines B and C are fitted by eye over the region within which Weber's law seems to hold.

where  $\sigma$  denotes intraindividual SD, and  $K$  and  $\psi$  category and magnitude scale values, respectively, and the straight lines

$$K = \alpha I + \beta, \quad (3)$$

were fitted by the method of least squares. Numerical integration was chosen throughout for the sake of compatibility, because explicit expressions could not be found for all Weber functions. As a general measure of goodness of fit, permitting comparisons between  $I_1$  (GPDE in general) and  $I_2$  (Fechnerian integration) as well as between all experiments, a normalized sum of squares was chosen, viz, stress, defined as

$$\sqrt{\frac{\sum (K' - K)^2}{\sum K^2}}, \quad (4)$$

where  $K'$  denotes the category scale values computed by Eq. 3, and  $K$ , the values obtained empirically.

The slope  $\alpha$  is of special interest. (The intercept  $\beta$  depends on the arbitrarily chosen values of the end points of the category scales and is thus of no interest.) If the model, as given by Eq. 1, is correct,  $\alpha$  should equal the "constant"  $\sigma_K(K)$  for Fechnerian integration ( $I_2$ ) and unity for GPDE ( $I_1$ ).

Disagreement between  $\alpha$  thus predicted and the  $\alpha$  obtained need, however, not entail rejection of the GPDE model.<sup>5</sup> If, otherwise, the fit of Eq. 3 is good, a discrepancy in  $\alpha$  means that the absolute sizes of the uncertainties as measured by the standard deviations are not quite comparable. There can be several reasons for this. The discrepancy might be real, in the sense that, e.g., the Os' set in the two experimental situations differs; there might also be computational artefacts, e.g., produced by calculating SDs over integers in the category scale and "continuous" values in the magnitude scale, or by the difference between the computation procedures for the two sets of SDs (see Eisler, 1962a).



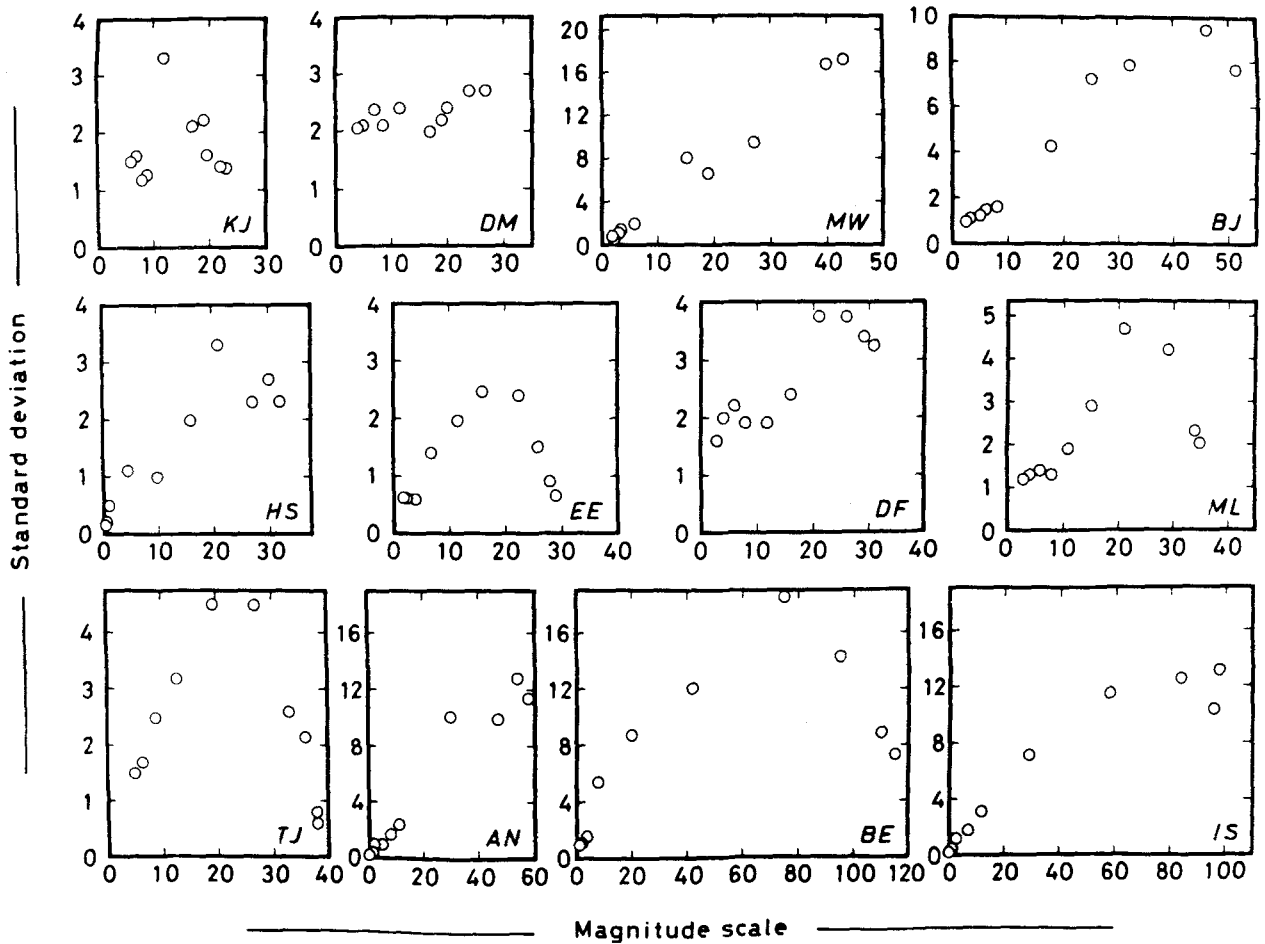


Fig. 11. SDs as a function of magnitude scales for loudness for individual Os. First row: ERA spacing (Experiment II); second row: ED spacing (Experiment IVa); third row: ED spacing (Experiment IVb). The Os are denoted by their initials.

**Results**

As Figs. 7 and 8 demonstrate, the fit of the empirical to the predicted category scales (Eq. 3) is good for all experiments, and often excellent, perhaps with the exception of O.H.S. It is noteworthy that the GPDE can cope with such peculiar relations between category and magnitude scale as those exhibited by Os E.E. and M.L. The fit can also be judged from the stress values in

Table 3, which vary between .012 and .08. For the 7-point category scales, better fit is obtained equally often for  $I_1$  and  $I_2$ ; for the 15-point category scale, however,  $I_1$  yields a better fit in all cases except one, where the fit is about the same for both methods. The small differences between the fit for the two methods,  $I_1$  and  $I_2$ , may seem surprising, since the deviations from constancy in the Weber function for the category scale

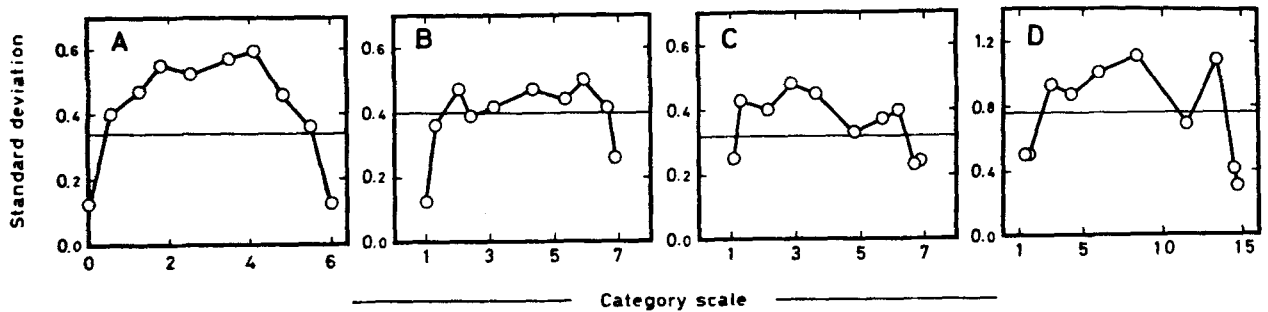


Fig. 12. Intraindividual SDs as a function of category scales for loudness. (A) Eisler's (1963a) previous experiment, (B) ERA spacing (Experiment I), (C) ED spacing (Experiment IIIa), (D) ED spacing (Experiment IIIb). The horizontal lines are predicted by Fechnerian integration.

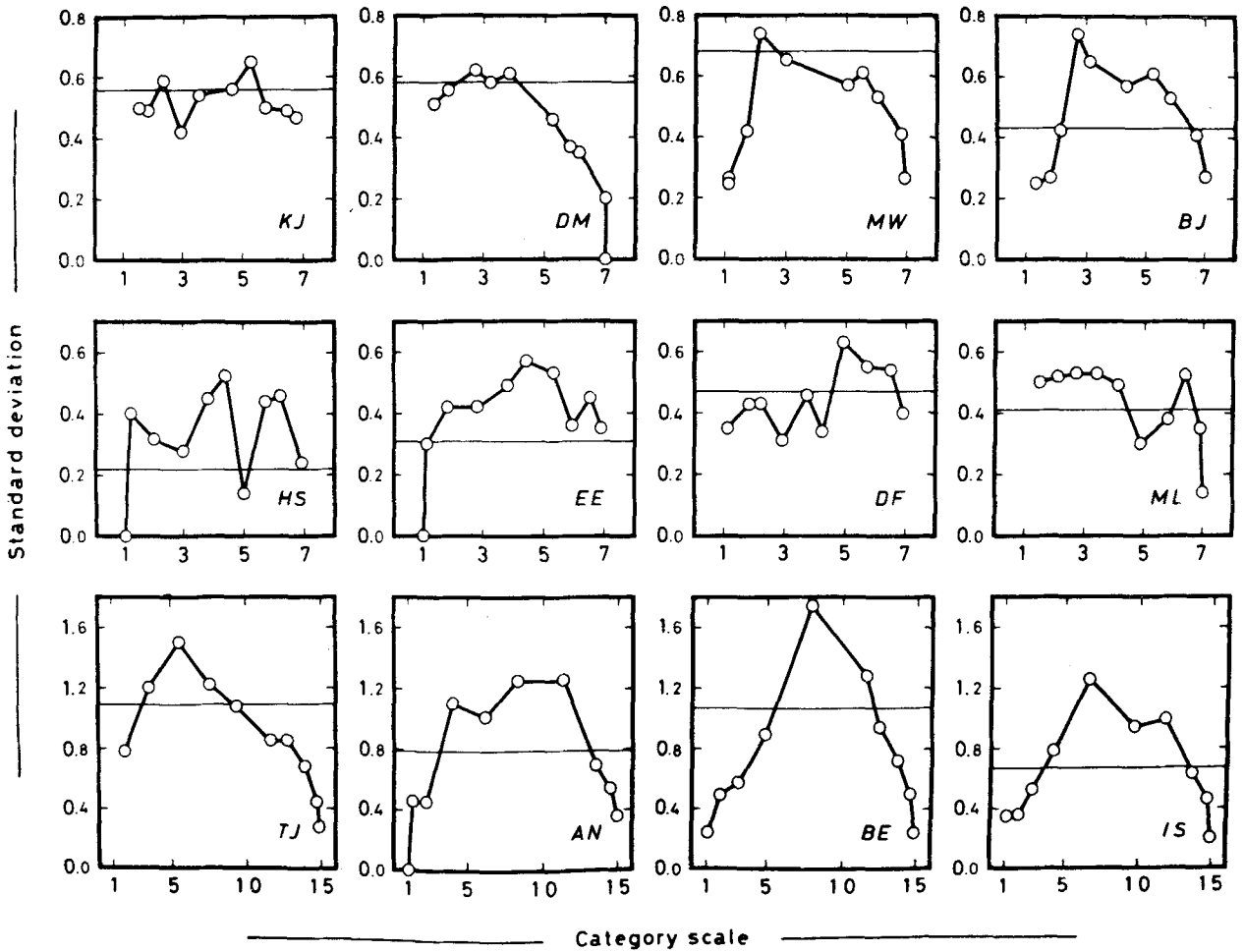


Fig. 13. SDs as a function of category scales for loudness for individual Os. First row: ERA spacing (Experiment II); second row: ED spacing (Experiment IVa); third row: ED spacing (Experiment IVb). As in Fig. 12, the horizontal lines are predicted by Fechnerian integration. The Os are denoted by their initials.

were occasionally rather large. However, those regions of  $\sigma_K(K)$  that deviated from constancy usually corresponded to relatively small portions of the magnitude scale, which implies that the deviations from a constant  $\sigma_K(K)$  had a relatively weak effect on Integral  $I_2$  in Eq. 2. A more sensitive test of the GPDE might be to compare, for each stimulus, the quantities on the two sides of Eq. 1, i.e., the derivatives of the category scale with respect to the magnitude scale to the ratios of the SDs. However, such a test requires reliable estimates of the derivatives, and these are difficult to find for the present data, since the function that relates the category scale to the magnitude scale is not known explicitly.

For the GPDE in general ( $I_1$ ), the slope  $\alpha$  should equal unity. For 13 out of 15 experiments, it is  $1.00 \pm .25$ . For two Os (H.S., E.E.), it is below .75. Too low an  $\alpha$  value indicates an overestimation of the uncertainty in the category ratings vs the magnitude estimations, and too high an  $\alpha$  value, the reverse (see Table 3). For the Fechnerian integration ( $I_2$ ),  $\alpha$  is the

constant uncertainty of the category scales. It is introduced as a straight line parallel with the K axis in Figs. 12 and 13 to allow a comparison with the empirical Weber functions.

### SUMMARY AND CONCLUSION

The attempt to obtain "ideal" stimulus spacings was successful for the ED spacing in the sense of an unambiguous convergence after a number of iterations. The equal-response-ambiguity (ERA) criterion, on the other hand, could only be approximated. The criterion is obviously labile; dips at either the extremes or the middle of the range could not be avoided (Fig. 2).

Figures 3 and 5 demonstrate that the magnitude scales obtained, and thus also the psychophysical relation, are strongly affected by the spacings. Whereas the group data for the ED spacing yield an approximately linear psychophysical function in log-log plots, following Stevens' power law, the function for the group data for the ERA spacing is markedly sigmoid.

The same trend is also found for the individual scales, though not so pronounced; none of the magnitude scales follows the power law perfectly. This observation is perhaps less surprising than the rather good agreement with the power law for the ED spacing's group data. J. C. Stevens (1958), for instance, has shown that crowding of the stimuli within a portion of the range investigated steepens the slope of the magnitude scale against stimulus values (in log-logs) in that portion, and both our "ideal" spacings crowded the stimuli towards the extremes of the range.

The plot of the 7-point vs the 15-point category scales (group data, ED spacing) is slightly concave downward. This indicates that the number of categories affects the form of scales obtained; an increase in the number of categories decreases the curvilinearity between magnitude and category scale (Montgomery, 1971).

Our first prediction failed in part. As Figs. 12B and 12C show, it proved impossible to get rid of the end effect in the Weber function of the category scales, though for both the ERA and the ED spacing, it was concentrated to Categories 1 and 7 for the 7-point category scales (group data); for the rest of the subjective range, the Weber functions were constant, with particularly small oscillations for the ERA spacing. The 15-point category scale (group data, ED spacing) showed a more gradual increase from the extremes towards the middle, Fig. 12D, though still weaker than for a set of stimuli with log spacing (Fig. 12A).

The second prediction, linearity of the Weber function of the magnitude scales including both end points, proved approximately true for the group data with ERA spacing. In contradistinction, group data with ED spacing exhibited a linear function only for the six lowest points of 10, including, however, the lowest one. Individual data were variable. It is thus seen that the ERA spacing is closest to the "ideal" case.

The third and last prediction, that the GPDE would permit calculations of the category scale from the magnitude scale and the two Weber functions, was fulfilled throughout, perhaps with the exception of the data from one O (H.S.). For group data, the relation between the two scales was the usual one: concave downward. For individual data, the form of the relation varied greatly, though it was never concave upward over the whole range. It is remarkable that the GPDE holds for the most irregular Weber functions, e.g., for the Weber function of the magnitude scale for the ED spacing. It should, however, be mentioned that in plots of the empirical vs the predicted category scales (not shown here), a slight sigmoid trend could be observed in several cases. This small discrepancy is obscured in the curves presented here.

The hypothesis that constant Weber functions for the category scales might give better predictions by the GPDE than empirical ones could not be confirmed (cf. Table 3). There was no clear difference for the 7-point category scales. This is in agreement with the aim of the

Table 3  
Slopes (Eq. 3) and Stress Values (Eq. 4) for Predicted Category Scales ( $I_1$ : GPDE in General,  $I_2$ : Fechnerian Integration)

Experiment	Slope ( $I_1$ )	Stress ( $I_1$ )	Stress ( $I_2$ )
I	.92	.0202	.0255
IIIa	.79	.0304	.0195
IIIb	.85	.0493	.0611
II (K.J.)	1.07	.0169	.0248
II (D.M.)	1.22	.0401	.0572
II (M.W.)	1.14	.0233	.0134
II (B.J.)	.76	.0655	.0462
IVa (H.S.)	.57	.0800	.0741
IVa (E.E.)	.69	.0200	.0120
IVa (D.F.)	1.06	.0505	.0559
IVa (M.L.)	.92	.0261	.0280
IVb (T.J.)	1.10	.0337	.0667
IVb (A.N.)	.80	.0554	.0728
IVb (B.E.)	1.05	.0704	.0692
IVb (I.S.)	.83	.0338	.0584

spacings to minimize the bias of the Weber functions, and the outcome is therefore not too surprising. For the 15-point category scale, on the other hand, the stress is lower when the empirical SDs enter into the computation. Perhaps one should conclude that the 15-point category scale is not a discrimination scale. Its almost parabolic Weber functions (see Figs. 12 and 13) are suggestive of the Weber function found in ratio estimation (Eisler, 1960). The fit for the 15-point category scale predicted by the GPDE is throughout not as good as it is for the 7-point category scales. One possible explanation may be that the unbiased Weber function is neither constant nor agrees with the empirical one.

The rather accurate predictions yielded by the GPDE taken together with the approximately constant Weber functions obtained for 7-point category scales may give some support for a distinction between the *genotypic*, i.e., unbiased, Weber function and the *phenotypic*, i.e., empirically obtained, Weber function of the category scale.<sup>6</sup> This is in agreement with the introductory remark that Weber functions are subject to bias due to experimental conditions and that this bias can be minimized with a suitable spacing of the stimuli.

## REFERENCES

- Attneave, F. *Applications of information theory to psychology*. New York: Henry Holt, 1959.
- Eisler, H. Similarity in the continuum of heaviness with some methodological and theoretical considerations. *Scandinavian Journal of Psychology*, 1960, 1, 69-81.
- Eisler, H. Empirical test of a model relating magnitude and category scales. *Scandinavian Journal of Psychology*, 1962a, 3, 88-96.
- Eisler, H. On the problem of category scales in psychophysics. *Scandinavian Journal of Psychology*, 1962b, 3, 81-87.
- Eisler, H. A general differential equation in psychophysics: Derivation and empirical test. *Scandinavian Journal of Psychology*, 1963a, 4, 265-272.
- Eisler, H. How prothetic is the continuum of smell? *Scandinavian Journal of Psychology*, 1963b, 4, 29-32.
- Eisler, H. Magnitude scales, category scales and Fechnerian integration. *Psychological Review*, 1963c, 70, 243-253.
- Eisler, H. On psychophysics in general and the general psychophysical differential equation in particular. *Scandinavian Journal of Psychology*, 1965a, 6, 85-102.
- Eisler, H. The connection between magnitude and discrimination

- scales and direct and indirect scaling methods. *Psychometrika*, 1965b, 30, 271-289.
- Eisler, H., Holm, S., & Montgomery, H. Is the general psychophysical differential equation an approximation? Report from the Psychological Laboratories, University of Stockholm, No. 386, 1973.
- Eisler, H., & Ottander, C. On the problem of hysteresis in psychophysics. *Journal of Experimental Psychology*, 1963, 65, 530-536.
- Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychological Review*, 1951, 58, 446-459.
- Helm, C. E., Messick, S., & Tucker, L. R. Psychological models for relating discrimination and magnitude estimation scales. *Psychological Review*, 1961, 68, 167-177.
- Luce, R. D., & Galanter, E. Discrimination. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- Montgomery, H. Direct estimation effects of methodological factors on scale type. Göteborg Psychological Reports, 1971, 1, No. 9.
- Siegel, W. Memory effects in the method of absolute judgment. *Journal of Experimental Psychology*, 1972, 94, 121-131.
- Siegel, J. A., & Siegel, W. Absolute judgment and paired-associate learning. *Psychological Review*, 1972, 79, 300-316.
- Stevens, J. C. Stimulus spacing and the judgment of loudness. *Journal of Experimental Psychology*, 1958, 56, 246-250.
- Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
2. The fit of the model underlying the equidiscriminability scaling procedure, which is essentially an application of Torgerson's (1958) law of categorical judgment (Condition D), appeared to be good. No systematic violations of the assumption of normal distributions could be seen, and the estimates of the discriminial dispersions tended to be constant.
3. Since in the individual experiments every O made 500 judgments, some effects of boredom or fatigue seemed quite probable. This possibility was investigated for the magnitude estimations of the four Os in Experiment II. For each stimulus, the 50 estimates were partitioned into blocks of 10 and the mean for each block plotted against block number. Apart from fluctuations in the absolute level of judgments (i.e., of the scale unit) no systematic trend could be discovered. The crests and troughs occurred at the same blocks for all stimuli and were roughly proportional in size to the scale values.
4. The peculiarity of some of the individual Weber functions obtained did not depend on the difference in the computing methods between group and individual data. This possibility was investigated by partitioning the 48 first judgments for each stimulus from two individual Os separately into 12 blocks of four judgments and, using the computation method for group data, by treating different blocks as different individuals. The correspondence between the SDs computed by these two methods was almost perfect for both Os.
5. The direction of regression is of no consequence for the data at hand. The greatest change in  $\alpha$  obtained for any of the 15 cases when the reverse regression line  $I = (1/\alpha)K + \beta$  was used amounted to a few percent.
6. Because of the very varied forms of the Weber functions of the magnitude scales, we refrained from arguing about genotypic Weber functions for those scales, identifying them tentatively with the phenotypic ones. This places the whole burden of distinction between phenotypic and genotypic Weber functions on the category scales.

## NOTES

1. A rigorous mathematical presentation of the general psychophysical differential equation is found in Eisler, Holm, and Montgomery (1973).

(Received for publication October 15, 1973;  
revision received March 14, 1974.)