VOL. 64, NO. 3 MAY, 1957

THE PSYCHOLOGICAL REVIEW

ON THE PSYCHOPHYSICAL LAW[1]

S. S. STEVENS

Harvard University

After dealing his critics their deserts, Fechner concluded his polemic of 1877 with a defiant five-line *Nachwort* (10).

> The tower of Babel was never finished because the workers could not reach an understanding on how they should build it; my psychophysical edifice will stand because the workers will never agree on how to tear it down.

For nearly a century the critics have pecked at the edifice in the disorganized manner that inspired no fear in Fechner. They have contended that sensation is not measurable, and that it is; that Weber's law is not true, and that it is; that the just noticeable difference is not a proper unit, and that it is; that the logarithmic law is not found in experiment, and that it is. As James put it, Fechner's critics flailed about, "smiting his theories hip and thigh and leaving not a stick of them standing . . . ," but here we are a hundred years later still discussing Fechner.

The lesson of history is that a bold and plausible theory that fills a scientific need is seldom broken by the impact of contrary facts and arguments. Only with an alternative theory can we hope to displace a defective one.

The purpose here is to try to do just that—to try to show that there is a general psychophysical law relating subjective magnitude to stimulus magnitude, and that this law is simply that equal stimulus ratios produce equal subjective ratios. On numerous perceptual continua, direct assessments of subjective magnitude seem to bear an orderly relation to the magnitude of the stimulus. To a fair first-order approximation, the ratio scales constructed by "direct" methods (as opposed to the indirect procedures of Fechner) are related to the stimulus by a power function of one degree or another. Evidence for this fact is piling up under the impact of ratio-scaling procedures whose development over the past three decades has given a "new look" to psychophysics. The fruits of these new pursuits have implications and consequences for the traditional issues in the area of psychophysics, as well as for certain related issues concerned with category judgments, time-order errors, and scaling theory.

That the Fechnerian philosophy of indirect measurement will fade quietly away as soon as it has been shown what direct methods can achieve is scarcely to be counted on, for elaborations and applications of this philosophy pervade important segments of the psychological activity. Curiously enough, modern Fechnerians are found less frequently among the psychophysicists than among the psychometricians and scale construc-

[1] This research was supported partly by a grant from the National Science Foundation, partly by Contract Nonr-1866 (15) with the Office of Naval Research (Project Nr142-201, Report PNR-188). Reproduction for any purpose of the U. S. Government is permitted.

tors. It is they who sometimes assume that "equally often noticed differences are equal." As elaborated in Thurstone's "law of comparative judgment," this principle is essentially Fechner's principle: namely, that the unit of measurement is given by resolving power. If not explicitly, at least by implication, this philosophy of indirect measurement asserts that all we can know about magnitude is what confusion tells us. Variability becomes the measure of things, and the mean is meaningless. But on an important class of those simpler psychological continua where these notions are testable, we can show that equally often noticed differences are *not* equal, and that a scale proportional to psychological magnitude is not achieved by procedures that try to transform variabilities, discriminal dispersions, or confusions into units of measure.

Before we consider this point further we will need to draw a distinction between two classes of continua.

Two Classes of Continua

Perceptual continua divide themselves into two general classes (64). The nature of this division is suggested in a general way by the traditional dichotomy between quantity and quality. Continua having to do with *how much* belong to what we have called Class I, or prothetic; continua having to do with *what kind* and *where* (position) belong to Class II, or metathetic. Class I seems to include, among other things, those continua on which discrimination is mediated by an additive or prothetic process at the physiological level (53, 67). An example is loudness, where we progress along the continuum by adding excitation to excitation. Class II includes continua on which discrimination is mediated by a physiological process that is substitutive, or metathetic. An example is pitch, where we progress along the continuum by substituting excitation for excitation, i.e., by changing the locus of excitation.

We know too little about physiological mechanisms to say whether all Class I continua are based on additive processes or all Class II continua on substitutive processes, but in those instances where the facts seem clear the parallels between function and physiology are at least suggestive. Until our knowledge stands improved, it is perhaps best to classify the perceptual continua by the pragmatic criteria of the way they behave and to hope that any uniformities we can discover will lead to deeper insights into basic mechanisms.

Four functional criteria are relevant to the distinction between prothetic and metathetic continua, although they have not all been tested with equal thoroughness. These four criteria concern the subjective size of the j.n.d., the form of category rating-scales, the time-order error, and hysteresis. It is likely, of course, that other criteria will be discovered if we continue to look for them.

1. *Subjective size of the j.n.d.* I list this criterion first because it is the growth in the subjective size of the j.n.d. as we go up the scale on a Class I or prothetic continuum that seems, in a sense, to "explain" at least two of the other three criteria. Fechner proposed that scales can be constructed by counting off just noticeable differences. The implication is that the sensation produced by a stimulus 50 j.n.d.'s above threshold is half as great as that produced by a stimulus 100 j.n.d.'s above threshold. The hard fact of the matter is that if the typical subject were confronted with two such stimuli on a Class I continuum he would assert with certainty that the ratio between the two sensations is greater than two, because scales obtained by summating j.n.d.'s are nonlinearly related to ratio scales of subjective magnitude (33, 52). But more about this in a later section. The

point here is that on Class I continua the j.n.d.'s are not equal in subjective size.

On continua of Class II (metathetic) the j.n.d.'s turn out to be approximately equal in subjective size when measured by magnitude scales of the continuum. The linearity between the mel scale of subjective pitch and the j.n.d. scale for frequency is a case in point (56) and it seems reasonable to suppose that the same linear relation is approximated on such continua as position and inclination.

2. *Category rating-scales.* A category rating-scale is the function obtained when a subject judges a set of stimuli in terms of a set of categories labeled either by numbers or by adjectives. The form of these scales is different on the two kinds of continua. As shown by studies of a dozen perceptual dimensions (64), the category scales on continua of Class I are concave downward when plotted against a ratio scale of the subjective magnitude. Category scales on continua of Class II *may be* linear when so plotted.

The chief factor that produces nonlinearity in the category scales of Class I is variation in the subject's sensitivity to differences. Near the lower end of the scale where discrimination is good the categories tend to be narrow, and by consequence the slope of the function is steep. Near the upper end, where a given stimulus difference is less easy to detect, the categories broaden and the slope declines. Only on Class II continua, where sensitivity (measured in subjective units) remains relatively constant, is it ordinarily possible to produce category scales that are linearly related to subjective magnitude.

Prothetic continua of Class I on which category scales have proved nonlinear include apparent length, area, numerousness, duration, heaviness, lightness, brightness and loudness. Metathetic

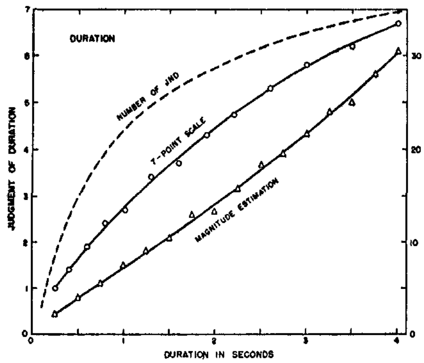


FIG. 1. J.n.d. scale, category scale, and ratio scale for apparent duration. *Triangles:* Mean judgments of 12 subjects who estimated the apparent duration of a white noise. Stimuli were presented in a different irregular order to each subject. *Circles:* Mean category judgments made by 16 subjects on a scale from 1 to 7. The end stimuli were presented at the outset to indicate the range, and each subject judged each duration twice in a random order. Stimulus spacing was adjusted to produce a "pure" category scale. *Dashed line:* Summated j.n.d.'s, right-hand ordinate.

continua of Class II on which category scales have proved more or less linear include visual position, inclination, proportion, and pitch. On both types of continua alterations in the form of the category scale may be produced by a variety of factors, including stimulus spacing (or relative frequency of presentation), landmarks, and differential familiarity.

By means of an iterative series of experiments we can achieve a stimulus spacing that will neutralize the effects of spacing and produce a "pure" category scale, uncontaminated by the subject's expectations regarding the frequency with which he should name the various categories. An example of an approximation to a pure category scale is shown in Fig. 1 (circles). This experiment was a follow-up to the experiments on duration reported earlier (64). Guided by the outcome of the earlier studies, I so spaced the stimuli that the 16 subjects tended to name the various categories equally often. The proof that a pure category scale has been approximated lies in the fact that if we use the curve in Fig. 1 to determine what spacing of the stimuli will mark off equal distances along the ordinate, no significant change in the spacing is called for. (The technique of experimental iteration is a powerful device for neutralizing certain kinds of bias. For other examples of its use see [57, 64, 66].)

The iterated experiments on duration culminating in that shown in Fig. 1 converge on a category scale that is typical of those obtained on continua of Class I (prothetic). We note that it is concave downward, as expected, and that it contrasts sharply with the ratio scale (triangles) determined by asking

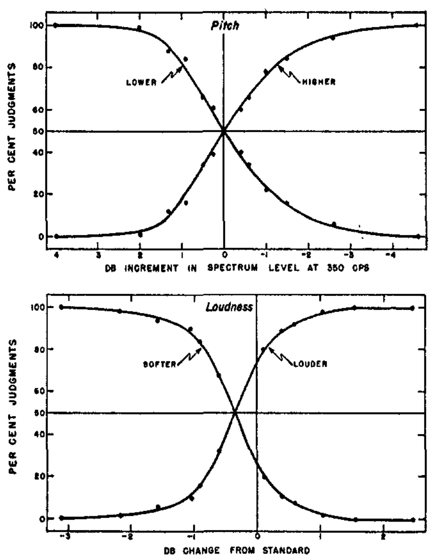


FIG. 2. Psychometric functions from 50 subjects obtained with recorded tests for ability to discriminate the pitch and loudness of bands of noise. Each test contained 110 test items consisting of a standard noise (2 seconds) followed immediately by a comparison noise (2 seconds). The band of noise used in the pitch test was shifted up or down the frequency scale. Note that there is a "time error" for loudness but none for pitch.

12 subjects to estimate directly the apparent duration of the stimulus. The upward curvature of the ratio scale for subjective duration agrees with the results of independent experiments in which the method of fractionation was used to determine the duration that appeared to be half as long as a standard duration (19, 50).

The curves in Fig. 1 tell us that on a prothetic continuum the typical subject is unable to equalize the intervals on his category scale, even when instructed to do so. He fails for the basic reason that his ability to tell one magnitude from another varies over the scale and affects the width of his categories. Since he can easily tell 0.5 sec. from 1.0 sec. he tends to put them in different categories; since he can only with difficulty tell 3.5 sec. from 4.0 sec. he tends to put them in the same category. Any procedure that tries directly or indirectly to get the subject to partition a prothetic continuum into equal intervals seems bound to fail in the general case.

For purposes of comparison, the function obtained by counting up j.n.d.'s is plotted in Fig. 1 as a dashed curve. This curve, based on Woodrow's (73) report that the relative j.n.d. for duration is of the order of 10 to 12.5 per cent, shows the form Fechner's law predicts for the scale of apparent duration.

3. *Time-order error.* The so-called "time-error" discovered by Fechner has been pursued in theory and experiment for many decades (23). It refers to the fact that the second of two equal stimuli tends to be judged greater than first. We have reason to believe that a systematic time-order error (as Guilford prefers to call it) is characteristic of judgments on Class I continua. On Class II continua we neither expect it nor do we generally find it.

This hypothesis was first suggested to me by the troubles we met when we tried to supply the armed forces with recorded tests designed to assess pitch and loudness discrimination for bands of noise (29). Karlin and I discovered that it was easy enough to record a pitch test whose scoring was straightforward, but how were we to score a loudness test when it always showed a time-order error? Were we to try to explain to the military that, although the second stimulus may actually have been 0.1 db lower than the first, if the soldier called it lower they should mark him wrong? We tried to get rid of the error by making the transition from the standard level to the variable level instantaneous, but the error was still there. We finally solved the issue by avoiding it: we dropped the stimulus items whose proper scoring was made ambiguous by the time-order error.

Figure 2 shows results obtained with these tests from groups of 50 subjects each. (In the course of this work several hundred subjects were tested.) The crossing points of the psychometric functions show that there is no time-order error for pitch, but that for loudness the error amounts to about 0.3 db. Similar results were obtained when we used the Seashore pure-tone tests. Postman (43) then took up the problem and confirmed our finding that there is typically a time-order error with loudness but not with pitch.

This perhaps is not the place to try to "explain" the time-order error, but I would like to venture an opinion about it. First of all, it is important to note that the error is typically small—a fraction of a j.n.d.—and hence is close to the limiting "noise level" of our measurements. It is therefore not surprising that the published findings resound with the noise of disagreement. Attempts to map the course of the effect as a function of time are particularly discordant, which is good reason for giving up the name "time-error." Its dependence on time is probably fortuitous. As a matter of fact, in an experiment still in progress (by G. M. Shickman) we have obtained the usual "time-error" when the subject lifts the standard and the comparison weight simultaneously, one with each hand.

What the Fechnerian time-order error really seems to depend on is the same basic process that makes the category scales on Class I (prothetic) continua turn out to be concave downward, namely, the asymmetry of sensitivity—the fact that discrimination is better toward the low than toward the high end of the range. The methods commonly used to measure the time-order error are essentially similar to those we use in category scaling. The subject places stimuli in two or more categories like *heavier, equal,* or *lighter* and the forces that determine his judgments are the same as in any category scaling procedure. (When the method of adjustment is used with loudness a different type of constant error occurs: the variable tends to be set *higher* than the standard [44], but the sign of the error may reverse when the standard is very loud [61].) The relation between the category scale and the conventional time-order error is perhaps best illustrated by means of Fig. 3, where we see three category scales for lifted weights covering different ranges of stimuli. (For a further discussion of these and other such curves, see [64].) The widest range (Cowdrick [7]) gives the usual category curve, concave downward, and we note that the middle weight in the series is assigned a value *above* the middle of the category scale. When the range is shortened (Guilford and Dingman [25]) the same general features are preserved. The shortest range is the one used by Fernberger (11) to determine a difference limen by the method of single stimuli. This curve, like the others, fails to pass

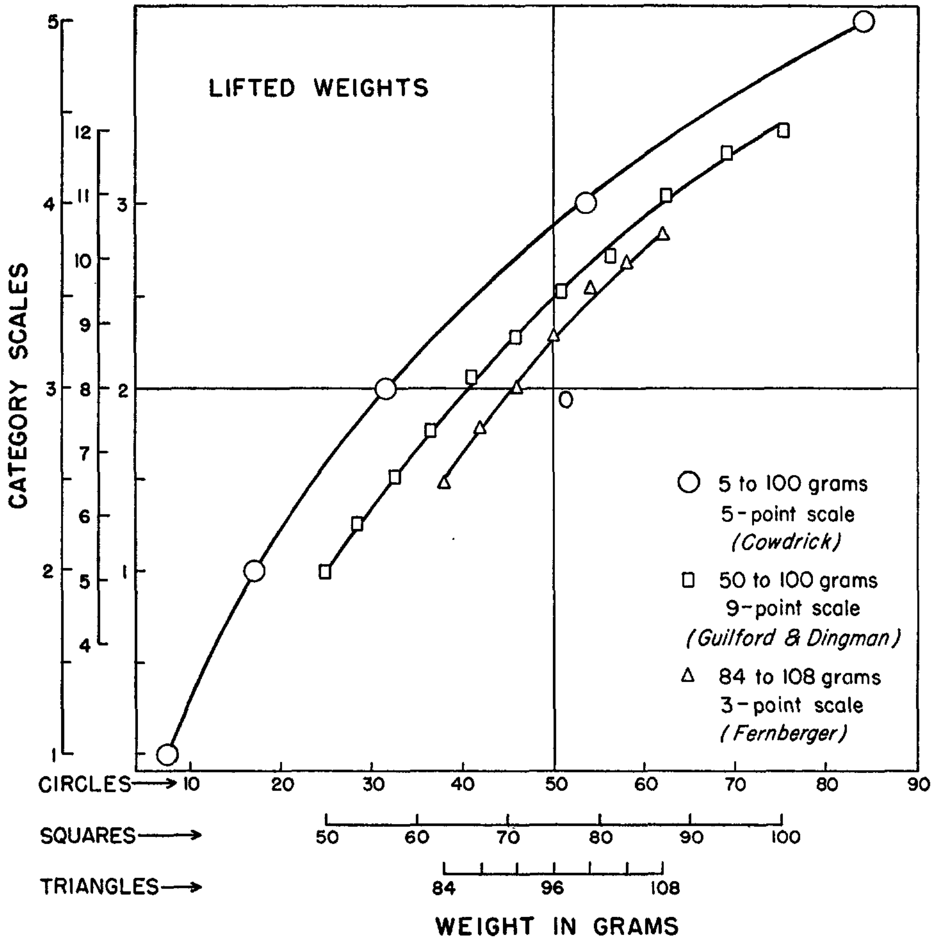


FIG. 3. Category scales for lifted weight. On the middle curve only every other point has been plotted. The end stimuli were designated 4 and 12 but the subject could use other categories if desired. Note that the curves are all concave downward. The fact that they do not pass through the point O indicates a "time error."

through the point O, which is the point defined by the middle of the stimulus range and the middle of the category scale. The distance between the point O and the point where the curves cross the horizontal line through the middle of the plot is the time-order error (cf. Pratt [46]). It is as simple as that. By means of the theory of category judgments we can apparently account for Fechner's time-order error without the help of such props as sinking traces, fading images, or assimilation. And it becomes clear why we should expect to find time-order errors on Class I continua (prothetic) but not on Class II (metathetic).

One further point. Not only is the time-order error produced by other factors than time, but it also has nothing to do with order. Nor is it necessarily an error. The name, like the one given to the Holy Roman Empire, is a misnomer in all particulars. If my thought is correct, that we are dealing here with an effect on category judgments that derives from an asymmetry based on the relativity of discrimination, a better name for it might be the "category effect." Such a label might also help to distinguish this form of bias from other sources of systematic "error," of which there are many. Not all constant deviations in psychophysical experiments can be blamed on the category effect. To take a single example, the observation of Fullerton and Cattell (13) that the second of two equal light flashes appears fainter than the first is probably due to an altered state of retinal adaptation produced by the first flash. In this instance an adaptation effect can apparently override the category effect and produce what would conventionally be called a positive time error. No doubt there are still other sources of systematic bias in experiments of this sort.

It is true of course that time may play a role in these and other systematic errors, in the sense that the effects may vary with the interval between judgments. But our problem is first to try to discover the basic nature of the asymmetries involved in the various kinds of constant errors and then to see how the forces that produce them may vary with time.

4. *Hysteresis.* Since this word means a lagging behind, as when magnetization lags behind the magnetizing current, it seems a good term to describe what happens when the apparent sense-distances between successive stimuli are judged in different orders. The effect shows up in especially dramatic fashion in experiments on bisection and equisection. In a typical experiment on loudness the subject sits before a row of five keys which he presses to produce the tones (1,000 cps). The levels produced by the two end keys are fixed, 40 db apart, and the subject adjusts the levels controlled by the intermediate keys in order to divide a 40-db interval into four equal-appearing steps in loudness.

Where the subject sets the levels depends on whether he listens to the loudnesses in ascending or descending order. I found that in the ascending order he sets the bisecting level some 5 to 8 db higher than in the descending order. In an experiment run to verify this effect, Garner (15) found an average discrepancy of 5.8 db. It is as though the loudness the subject hears lags behind what he should hear as he goes up or down the scale. As a result of this lag the graphs of the functions shown in Fig. 4 exhibit a "hysteresis loop": the ascending path is different from the descending path. Please note, however, that in calling this phenomenon hysteresis I am trying to describe it, not explain it. I am not sure I know how to explain it.

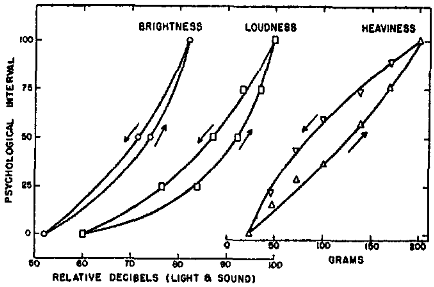


FIG. 4. Hysteresis effects in judgments of intervals on three sensory continua. The arrows indicate the order of stimulus presentation. For brightness and loudness the subject adjusted the middle stimuli to divide the range into two or four equal-appearing intervals. For heaviness the subject lifted three weights and indicated the relative apparent position of the middle weight by means of a slider between two markers on a bar. Five different middle weights were used in ascending and descending order.

As shown in Fig. 4 the same hysteretic effect occurs in the bisection of brightness intervals and in judgments of the intervals between lifted weights. Partly because the range was shorter (30 db) the brightness bisections show less hysteresis than loudness. Lifted weights, on the other hand, show relatively more. The possibility suggests itself that the magnitude of the hysteresis may vary inversely with differential sensitivity. In judging the apparent size of the two intervals between three weights, the subjects moved the middle of three pointers on a steel bar to indicate the relative position of the middle weight. Thus we see that hysteresis is independent of whether the subject himself adjusts the stimulus. The same bar and pointers were also used for judgments of loudness, and the resulting hysteresis was essentially similar to that shown in Fig. 4. The data in Fig. 4 are typical examples from an extensive series of unpublished experiments, some of which were run by R. J. Herrnstein.

Hysteresis occurs on the three examples of Class I continua that we have tested. The hypothesis I am suggesting is that it may occur on all continua of this class, but that it probably does not occur on continua of Class II. Evidence for this possibility is suggestive, but not yet conclusive.

Bisections on the continuum of visual position, for example, seem to be free of hysteresis. At least what minor effect has been found is in the opposite direction from that observed on Class I continua. Volkmann kindly undertook to test this matter on the large viewing screen (21.6 × 6.8 ft.) at Mt. Holyoke College (6). The subject sat 35 feet from the screen and adjusted the position of the middle of three points of light that came on in the sequences right-middle-left and left-middle-right. The end lights were a meter apart. When the 12 observers were instructed to fixate the center of the screen, the two orders produced an average difference in the bisection point of only 0.25 cm. When the instructions were to look at each stimulus as it came on, the discrepancy was 3.6 cm. Since both discrepancies were in the opposite direction from those observed with loudness, brightness, and lifted weights, we may conclude that, whatever constant errors may occur on this Class II continuum, they apparently constitute a different phenomenon from the Class I hysteresis.

The question of hysteresis in pitch judgments is less easy to answer. We found no evidence for hysteresis in the experiment that led to the construction of the mel scale (67), but the fact is we were not looking for it. In a more direct test of the issue I later had 10 subjects bisect various pitch intervals in ascending and descending order. The experiment was casual and not definitive, and the results were not clear cut. For some subjects hysteresis seemed to be present, for others not. Two subjects happened to have "absolute pitch"—the best cases I happen to have encountered. Each of these two subjects set the middle tone to the same frequency on every trial, regardless of the order of listening. So it seems that at least some people show no hysteresis in their pitch judgments. But this is apparently not true of others.

In a recent laboratory exercise, two of my students, D. D. Greenwood and M. L. Israel, found hysteresis in the equisections of all but one of nine subjects who divided the interval 400–7,000 cps into four equal-appearing intervals.

Incidental evidence for the absence of hysteresis in pitch bisections is found in an experiment by Cohen, Hansel, and Sylvester (5). They were concerned with another problem, but part of their procedure called for the bisection of the pitch distances between 1,000 and 3,000 cps and between 2,000 and 4,000 cps in both ascending and descending order. The bisections were as follows:

For 1,000–3,000 cps, going up: 1,874 cps; going down: 1,838 cps.

For 2,000–4,000 cps, going up: 2,693 cps; going down: 2,808 cps.

The average standard deviation for all settings was 303 cps.

We see that the interval 1,000–3,000 cps shows an insignificant discrepancy (36 cps) and that the larger discrepancy

(115 cps) shown by the interval 2,000–4,000 cps is less than half the standard deviation. It is also important to note that the directions of the discrepancies between the results for ascending and descending orders are opposite in the two parts of the experiment. Incidentally, the points of bisection that are predicted by the mel scale are 1,859 cps and 2,280 cps, which are within 3 per cent of the observed values.

All in all then, the evidence for hysteresis in pitch bisections is ambiguous. The pitch continuum probably means different things to different people, especially to those with and without "absolute pitch." It is in many ways a difficult continuum about which to generalize (see, for example [64]).

Another point concerning bisection deserves mention. On continua of Class I (prothetic) the point of bisection falls consistently below the point predicted by the ratio scale of the subjective magnitude (39, 58, 63). This is true of the combined results from which hysteresis has been eliminated by an averaging of the ascending and descending judgments. It is also clearly evident in the early experiments by Fullerton and Cattell (13) who explored the "force of movement" by the methods of halving, doubling, and bisection. This error in bisection is scarcely surprising when we consider that bisection and category rating-scales have much in common. In both instances the subject tries to equalize intervals, and in both instances he misses the mark in the same direction—presumably because both kinds of judgments are controlled partly by discrimination, and discrimination varies from one end of the continuum to the other.

On continua of Class II (metathetic), on the other hand, the results of bisection agree with those obtained by fractionation and by magnitude estimation. This has been demonstrated for pitch (64, 67) as well as for inclination or "bearing" (47, 64). Apparently it is only on Class II continua that the construction of a scale of subjective magnitude by the method of bisection leads to a valid outcome, for only there can bisection be shown to agree with direct magnitude estimation.

In our experiments on weight and loudness in which the subject moved a pointer along a bar, the parallel with category rating-scales is even more obvious. In effect, the bar and pointer constitute a continuous rating scale, analogous to the widely used scales on which the rater marks a point on a line with a pencil. The bar and pointer also give results (hysteresis averaged out) that differ from those predicted by the ratio scale of the Class I continuum.

We see then that possibly four functional criteria can be used to divide the perceptual continua into two general classes. Clear examples of each of these classes are easy to find, but this is not to say that all examples one might think of will prove easy to classify. When we try to impose a simple order on the complex fabric of our experience there is an ever-present risk that we will oversimplify what is not simple. Class I (prothetic) constitutes a rather unitary and well behaved group of continua, but Class II (metathetic) includes various disparate aspects of things, many of which do not fall naturally onto unitary monotonic scales, and some of which may not even be orderable. Some continua may turn out to be hybrid mixtures of both Class I and II. Pitch, in fact, may be one of these (64).

In most of what follows we will limit our concern to the prothetic continua (Class I), the "quantitative" aspects of things, for it is here that we can most profitably examine the principles that psychophysics has set itself to devise. The first of these principles concerns the relation between stimulus and sensation.

THE POWER FUNCTION

There is a growing bundle of evidence to indicate that on prothetic continua the form of the "psychophysical law" is a power function. Most of this evidence, accumulated since 1930, is the product of a reoriented psychophysics that brings "direct" methods to bear on the Fechnerian query regarding the relation between stimulus and sensation. These direct methods lead to ratio scales of perceptual magnitude, and to a first approximation these scales show that the psychological magnitude is a power function of the stimulus magnitude.

The basic principle seems to be that equal stimulus ratios tend to produce equal sensation ratios. Empirical evidence to support this principle has been obtained for at least a dozen perceptual continua, with the aid of several different experimental methods. What the principle entails can be illustrated by examples such as these: In order to make one sound seem half as loud as another, the physical energy must be reduced by about 90 per cent (10 db), and this required reduction is approximately the same regardless of the level from which we start. The same is true of brightness, and the required reduction is of the same order of magnitude as for loudness. In order to make one lifted weight seem half as heavy as another, the original weight must be reduced by about 38 per cent, and this percentage reduction is approximately constant over a wide range of stimuli. The approximate constancy of the percentage reduction corresponding to a given subjective ratio demonstrates that the sensation ψ is proportional to the stimulus S raised to a power n. Thus

$$\psi = kS^n.$$

When we convert this equation to logarithmic form we obtain a linear equation which has a certain practical usefulness, for the function can then be represented by a straight line in log-log coordinates. The slope of this line corresponds to the exponent n.

When we know the value of ratio r between the stimulus judged half and the standard stimulus we can obtain the value of n by dividing log r into log 0.5. In general, if s is the sensation ratio (determined by the method of ratio production) and r is the corresponding stimulus ratio, then

$$n = \frac{\log s}{\log r}.$$

It is not to be expected, of course, that a simple power function will hold from zero to infinity. What the limits may be must be decided by experiment. We know that some functions relating stimulus to response exhibit discontinuities (e.g., apparent numerousness and flash rate), and we know that our formulas might need eventually to take account of second-order perturbations in the vicinity of the absolute threshold (cf. 8, 58). But our present concern is with first-order approximations to general laws.

METHODS FOR RATIO SCALING

Before we consider the concrete instances of ratio scales, let us look briefly at the problem of procedure. Methods for constructing ratio scales of subjective magnitude are a relatively new development. It is true that Merkel (31) with his *Methode der doppelten Reize* (1888) tried to find the stimulus that appeared to double the sensation, but Merkel's effort seems to have had little influence on psychophysics. Fullerton and Cattell (13), who used the methods of doubling and halving, fared little better. Even Titchener (72) in his thorough and scholarly Instructor's Manual makes only off-hand allusion to Merkel's method of "doubled stimuli." And of the author of the method he says "we are reminded that he is, by mental constitution, a physicist rather than a psychologist . . . and that his work as a whole is exceedingly poor in introspective data" (p. 223). It may be that "mental constitution" plays a role in these things, as Titchener suggests. I wonder what he would make of the fact that at least seven different physical laboratories have made important contributions to the development of the ratio scale of subjective loudness—the sone scale—as against perhaps three different psychological laboratories.

The impetus for the development of a method is a problem. Method for its own sake, tempting and fascinating as it may be, often leads to little—except methodology. Physicists and psychologists went to work to refine the procedures for measuring loudness mainly because the substantive outcome was of interest and importance to someone, particularly to acoustical engineers. This is shown by the fact that some of the earliest studies were paid for by commercial companies. Curiously enough the practical problem originated in an obvious failure of Fechner's law. Not long after they had adopted the decibel scale for measuring sound intensities, the engineers noted that equal steps on the logarithmic (decibel) scale do not behave like equal steps, for a level 50 db above threshold does not sound at all like half of 100 db, as Fechner's law implies it should. Since the acoustical engineer must often interpret to his customers the meaning of esoteric acoustical measurements, it soon became clear that a scale was needed on which the numbers would be proportional to how loud things sound to the typical listener. Without the motive of this kind of practical problem to spark the development of method, it is a fair guess that the ratio scales for the fourteen perceptual continua we will examine might never have been constructed. How a subjective ratio scale, once achieved, can be put to work is shown by the way we were able to use the sone scale in the development of a procedure for calculating the loudness of a complex noise from a spectral analysis of the sound (61).

The methods for constructing ratio scales are still in the process of development, but in one form or another they all require subjects to make quantitative estimates of subjective events. Many authors have screamed that this is nonsensical, meaningless, and impossible (12, 34), but those who follow these methods go ahead and do it anyhow. These direct assessments of sensation seem not so impossible after they have been made.

At present the methods are principally four, but each has subvarieties. We may classify the methods more or less systematically as follows:

1. Ratio estimation
   a. Direct judgment of ratios
   b. "Constant-sum"
2. Ratio production
   a. Fractionation
   b. Multiplication
3. Magnitude estimation
   a. Prescribed modulus
   b. No designated modulus
4. Magnitude production

I am well aware that few activities in psychology exhibit greater terminological confusion than the naming of methods, and that there is probably no effective cure for the burgeoning, ramifying process that occasions the welter of names. The best we can do is take stock from time to time and attempt a little systemization. Having formerly contributed my full share to the confusion, in the foregoing list I have tried for a dose of order.

Partly through my efforts the second class of procedures, which was the first used historically, got the name fractionation (68), because the usual procedure has been to require a subject to set one stimulus to produce a sensation half as great as a given standard. Frac-

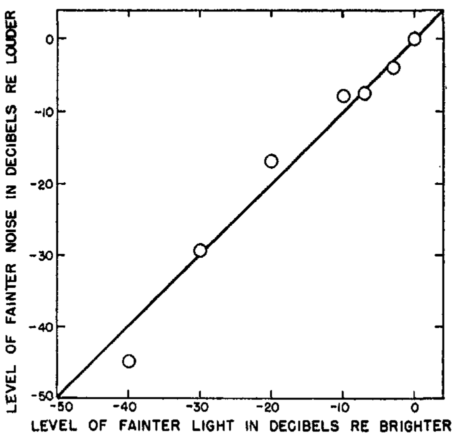


FIG. 5. Results of adjusting a loudness ratio to equal an apparent brightness ratio. The subject viewed a pair of luminous circles of which one was dimmed by an amount shown on the abscissa. He also pressed a pair of keys to produce two different levels of white noise. He adjusted one level (ordinate) to make the loudness ratio seem equal to the brightness ratio. The brighter light was about 99 db re 10^{-10} lambert and the louder noise was about 92 db re 0.0002 microbar.

tions other than a half are also used with consistent results. But fractionation in this sense is only one facet of a more general procedure. The other facet, which we might call *multiplication,* involves the complementary process of requiring the subject to identify or produce a prescribed ratio that is greater than one, e.g., to double, triple, etc., a given standard. This procedure has not been used very much, probably not as much as it should be, for there is abundant evidence (62) to indicate that, when it is made to complement halving, the procedure of doubling can help to balance out certain systematic biases.

The two procedures together may be called *ratio production.* The production of ratios can be carried out by a variety of different techniques. Thus the experimenter may allow the subject to adjust a stimulus to produce a prescribed ratio to a standard, or the experimenter may set the stimulus and ask the subject whether it meets the prescribed ratio (method of "constant stimuli").

An interesting variant of ratio production involves the fixing of two luminances to define an apparent ratio which the subject is to reproduce by setting two loudnesses to the same apparent ratio. In an experiment by J. C. Stevens in this laboratory the subjects set nearly the same *physical* ratio (decibels) between the intensities of the white noises as the experimenter set between the intensities of the two white surfaces. Figure 5 shows the median settings made by 15 subjects, each of whom twice adjusted the loudness of the second of two noises until the apparent ratio between the noises equaled the apparent ratio between two luminous targets viewed against a dark surround. The subject sat about 4 feet from the targets, which were 4.4 cm. in diameter and separated by about 17.5 cm. The fact that the median settings of the variable noise fall fairly close to the line indicates that the subjective impressions produced by white light and white noise are similar functions of stimulus intensity. The adjustments made to the largest ratio (40 db) were of course quite variable: the semi-interquartile range was about 10 db.[2]

The method of *ratio estimation* is the inverse of ratio production. Instead of prescribing the ratio in advance, the experimenter presents two (or more) stimuli and asks the subject to name the ratio between them. The subject may name the ratio directly, as in the pioneer experiment of Richardson and Ross (49), or he may be constrained to express the ratio by dividing a given number of points between the two stimuli in the manner proposed by Metfessel (32). The constraint involved in the so-called "constant-sum" method has

[2] For the rationale of measuring luminance in decibels see (59).

obvious disadvantages when large ratios are involved.

The method of *magnitude estimation* (60) dispenses with ratios as such and requires the subject to assign numbers to a series of stimuli under the instruction to make the numbers proportional to the apparent magnitudes of the sensations produced. The experimenter may prescribe a modulus by presenting a given stimulus and calling it some particular value, e.g., 10, or he may leave the subject free to pick his own modulus. (Note: if the magnitude estimations give a skewed distribution, as they usually do, it is advisable to calculate medians instead of means.)

The method of magnitude estimation has a logical inverse in *magnitude production*—a possible method that has thus far been mostly ignored. Instead of presenting a series of stimuli in irregular order and asking the subject to judge their apparent magnitudes, the experimenter might name various magnitudes and ask the subject to adjust stimuli to produce proportionate subjective values. Like any given procedure, this method probably has its "bugs" and biases and it might be interesting to find out what some of them amount to.

One thing we know is that in using this method the experimenter must resist whatever impulse he may have to fix or designate the top and bottom of the range, for this will change the problem into one involving category scaling. In some earlier experiments (64) we used this procedure, which we may call category production, to generate a 7-point category scale of loudness. We presented two levels which we designated 1 and 7 and then asked subjects to produce each of the other categories in irregular order. The results are like the usual category judgments obtained on Class I continua: the function is concave downward when plotted against the sone scale.

Magnitude production of a sort has been used in an experiment in which we used a brightness rather than a number to designate a magnitude. We set the luminance of a single target to various levels and asked subjects to adjust a noise to make its loudness appear as great as the brightness of the light. Although this investigation is still in progress (by J. C. Stevens) the results appear to be fairly consistent with what we know about the subjective scales for loudness and brightness. The intensities of the white noise are made roughly proportional to the intensities of the white light, again suggesting that loudness and brightness are similar functions of intensity.

All four of the methods listed above provide the kind of data necessary for the construction of a ratio scale (55). Each method can of course be varied and modified in numerous ways. Not only do the methods need to be altered and adapted to fit the problem at hand, but in any serious effort to establish a definitive scale for a given perceptual continuum it is desirable to ferret out the possible sources of bias by using different methods and by altering relevant parameters. In the present state of the art a valid scale that is representative of the typical subject can scarcely be hoped for from a single try.

The following ratio scales of subjective magnitude have been erected by means of one or more of the general procedures named above. They have not all been given the intensive study and cross-checking they deserve, but our interest here is in their general form rather than in their detailed adequacy. To a first approximation they are all power functions.

FOURTEEN RATIO SCALES

Our review of the fourteen examples of ratio scales of subjective magnitude will be brief, because most of the details

TABLE 1

PERCEPTUAL CONTINUA ON WHICH PSYCHOLOGICAL MAGNITUDE IS A POWER FUNCTION OF THE STIMULUS

The second column shows the approximate exponent of the power function. Names suggested for the various subjective units are listed in Column 3. The methods used (Column 4) are indicated by numbers that refer to the methods listed elsewhere in this paper.

Continuum	Exponent	Name of Unit	Methods Used	Reference Source
Loudness	0.3	sone	la, 2a, 2b, 3a, 3b, 4	(58)
Brightness	0.3–0.5	bril	2a, 2b, 3a, 3b	(28), (64)
Visual distance	0.67		2a	(17)
Taste	1.0	gust	2a	(1)
Visual length	1.1	mak	1a, 2a, 3a	(8), (47), (64)
Visual area	0.9–1.15	var	1a, 1b, 2a	(8), (47)
Duration	1.05–1.2	chron	2a, 3a	(19), (50), 64)
Lightness of grays	1.2		3a, 3b	(64)
Finger span	1.3		3a, 3b	*
Numerousness	1.34	numer	2a	(69)
Heaviness	1.45	veg	1a, 1b, 2a, 2b, 3a, 3b	(64)
Auditory flutter	1.7	flut	2a, 2b	(42)
Visual velocity	1.77		2a	(9)
Visual flash rate	2.0		2a	(48)

* S. S. Stevens and Geraldine Stone. Research in progress.

are available elsewhere. We can help the cause of brevity by means of Table 1 where the significant facts are listed and the sources of more detailed information are tabulated. In Table 1 the continua are ordered by increasing size of the exponent of the power function, and whenever a name for the subjective unit of the ratio scale has been suggested it is recorded. The methods that have been applied to each scale are indicated by numbers that refer to the listing in the previous section. I think all these continua belong to Class I, but this of course may be open to question.

Since Table 1 was prepared, an experiment on smell has been started in this laboratory by T. S. Reese. The early results suggest that the apparent intensity of the odor of benzaldehyde (synthetic almond) is a power function of the concentration in moles per liter of air. The exponent seems to be of the order of 0.2.

It goes without saying that all these scales have not been determined with equal care and thoroughness. About some of them we know a lot and have reason for confidence in their form; others have received only limited attention. Some of these attributes are easy to scale; others are more resistant to precision. Some of them represent perceptions about which we frequently make judgments of a more or less quantitative sort and about which the subjects in an experiment know, or think they know, so much about the stimulus that the "stimulus error" (3) may be a ready danger; others involve stimuli whose physical parameters are so unfamiliar to the typical subject that his temptation to try to estimate a physical measure of the stimulus, rather than to judge how it appears, is remote and uncompelling. Yet all these ratio scales approximate power functions.

I say approximate because it is certain that not all judgmental continua are strictly power functions of an arbitrary physical parameter of the stimulus. We know, for example, that over most of the range the loudnesses of a 1,000-cycle tone and a white noise grow with intensity according to similar laws, but at very low intensities the loudness of the white noise grows more rapidly. Hence the loudnesses of noise and tone cannot both follow power functions throughout their entire range, and it is even doubtful that either is strictly a power function at better than a first-order approximation (62).

On this question of the form of a psychophysical law we ought perhaps to try to learn a lesson from Fechner's mistake and forego the temptation to insist on the exactitude of a particular formula. In the realm of human judgment we may aspire to the discovery of first-order invariances, but it would be idle to pretend that precision can be pushed to the level envisaged by Fechner. The judgment of subjective magnitude is inherently a "noisy" phenomenon. When people try to describe a sensation in quantitative terms they face a difficult task, and the factors that affect the outcome are numerous and subtle. Patience and experimental skill can probably clean up part of the variance, but there will always remain irreducible dispersions to set a level below which we sink into uncertainty. In the broader perspective of things the power law stands out as a first-order relation, just as Weber's law describes the first-order relation between magnitude and resolving power. Although we are sometimes more fascinated by second-order departures from first-order relations than we are by first-order relations themselves, the first-order generalities have the broader significance. It is of greater moment to know the first-order law for the velocity of falling bodies than it is to prove that a host of variables can produce departures from the law.

Brief comments on some of the scales listed in Table 1 are now in order. We have already discussed loudness in various connections, and a review of most of the work on the problem is available (58). An example of a segment of the loudness function obtained from 32 subjects by the method of magnitude estimation is shown in Fig. 6. It is interesting that the sone scale is, so far as I know, the only subjective ratio scale for which the American Standards Association has appointed a committee to recommend a standard form.

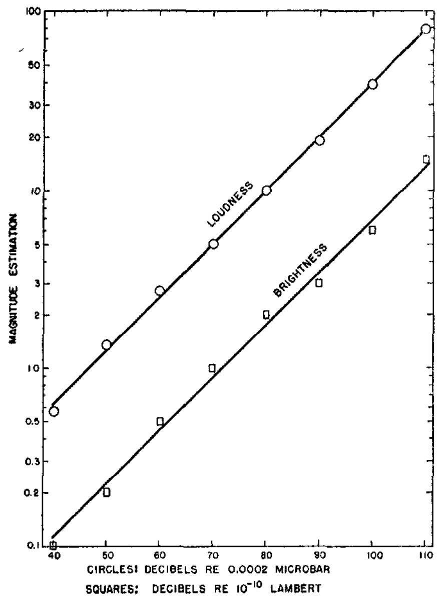


FIG. 6. Median magnitude estimations for loudness and brightness. For loudness each of 32 subjects made two estimates of each level (1,000 cps) in irregular order with no designated standard. Estimates were transformed to a common modulus at the 80-db level. For brightness each of 28 dark-adapted subjects made two estimates of each level in irregular order. The target subtended about 5° and was illuminated for about 3 sec. at each presentation. Once at the beginning of each session the subject was shown the level 70 db (14 subjects) or 80 db (14 subjects) and told to call that level 10. Estimates for all subjects were transformed to a common modulus at 70 db.

Brightness is still under active study. For ordinary luminous targets seen against a dark surround by a dark-adapted observer, brightness seems to grow approximately as the 0.3 power of the luminance. Figure 6 shows magnitude estimations for brightness obtained from 28 subjects in an experiment being conducted by J. C. Stevens. For very small targets that approximate point sources the exponent is larger than that determined by the line in Fig. 6, and of course the state of adaptation of the eye and the presence or absence of contrast makes a difference to the brightness function. These two factors increase the size of the exponent, as can be seen in Table 1 where the exponent for lightness of gray papers is listed as 1.2. When grays of different reflectances are viewed under ordinary illumination, light adaptation and contrast both operate to increase the steepness of the brightness function to a different order of magnitude from that observed with luminous targets seen in the dark.

Visual distance refers here to the apparent distance from the observer to a marker on a 200-foot line that stretched away from the observer on a large flat lawn. Two subjects made fractionation judgments. The combined results can be fitted by a $\frac{2}{3}$ power law as well as, if not better than, by the formula proposed by Gilinsky (17). Specifically, the data in her Fig. 6 fall close to a straight line. The main difficulty with the fit occurs at the shorter distances, where for one observer the apparent half values are nearly equal to the objective half values. But we would expect apparent distance to approximate objective distance when the angle between the line on the grass and the observer's line of regard becomes large. Perhaps if the line went out from the bridge of the observer's nose we would obtain a power function all the way—with an exponent less than one. Of course apparent distance, like apparent size, is specific to the experimental situation and is affected by cues of various sorts. It is not to be expected, therefore, that a single power function will govern all judgments of distance. As a matter of fact, in a very different set-up involving short distances (4.5 meters and less) Gruber (21) obtained data that fit a power function with an exponent of approximately 1.3.

Taste, curiously enough, gives an exponent of approximately unity when the stimulus measure is the ratio of solute to solvent (1). The functions reported for four different taste substances vary somewhat from one to another, for taste is a troublesome modality to work with, but they all give fairly straight lines in a log-log plot and the slopes hover around the value one. Another experiment (on sweetness [30]) gave only a rather poor approximation to a power function.

Length and duration give subjective ratio scales that are nearly proportional to the physical scales on which we usually measure the stimulus. On balance, however, there is a tendency in both cases for the exponent of the power function to be greater than one, which means that on these continua the magnitude of the subjective impression is a slightly accelerating function of the stimulus magnitude. The accelerating function for duration is shown in Fig. 1.

Apparent area seems to be nearly proportional to physical area, but the shape and size of the area to be judged seem to produce second-order differences. For smaller squares the exponent appears to be greater than for larger squares. In a recent experiment with circles the exponent turned out to be less than one (8).

For continua such as length it is sometimes held that learning plays a major role: we learn what length measures half as long as another, and so it comes to look half as long—which is one way of explaining why the exponent of the power function is near unity. This may or may not be true. Of course, if by learning we mean that the organism has had a history, we can never hope to experiment on subjects who have not learned. How the organism got to be the way it is will always be an engaging problem, but it is also a worthy object of attention to discover just how the typical subject now behaves, regardless of how he got that way.

The "learning explanation" of these ratio scales seems much less convincing for such continua as loudness and brightness. Despite Ebbinghaus' argument (see [60]) that we learn the meaning of a doubled brightness by illuminating the scene with two lights instead of one, the typical viewer actually requires some nine or ten times the illumination before he reports the brightness doubled. Why is this? And why is it that when the reflectance of gray papers is involved, where it is scarcely conceivable that the typical viewer knows the nature of the physical processes at work in the stimulus, the exponent is greater instead of much less than unity? Sophistication as regards the stimulus seems less than a likely explanation.

It is true, however, that one can learn to estimate the physical magnitude of a stimulus, as a good photographer well knows. He corrects what he perceives by what he knows about light and its relation to subjective impressions. By a similar procedure the acoustical engineer may estimate decibel levels with fair precision. He learns what a given subjective impression means in terms of the indications on a level meter. I have tested at least a half dozen acoustical engineers whose business involves the estimation and measurement of decibel levels. Despite their familiarity with decibels their subjective scales of loudness are not systematically different from those of other listeners. They can tell me, for example, both that one sound seems about half as loud as another, and also that the one is about 90 db and the other about 100 db.

Let us return now to Table 1. The ratio scale for finger span was obtained as part of a program that is still in progress. Subjects judged the apparent thickness of blocks of wood placed for about a second between the thumb and middle finger. The magnitude estimations obtained show that over a range of from about 0.2 to 6.0 cm the impression of felt thickness grows as a power function of the thickness of the blocks, and that the exponent is about 1.3.

Numerousness refers to the impression one gets from looking at a collection of dots that contains too many to be counted. The exponent of about 1.34 was derived from Taves' experiment in which he used the method of fractionation. I computed the exponent from the mean fraction judged half for all standards above 8 dots. The exponent is equal to log 0.5 divided by the logarithm of the fraction judged half.

Lifted weights have been scaled by several methods in several laboratories and most experimenters have found close approximations to power functions. The value of 1.45 for the exponent of the veg scale was obtained from pooled results from five different laboratories. For the procedures involved, see (64). It is interesting to note that when Fullerton and Cattell in 1892 had their subjects try to halve and double the apparent force exerted on a dynamometer they came quite close to the values that would be predicted by the veg scale. The two values obtained fell on opposite sides of the predicted value, and the largest discrepancy was about 10 per cent of the predicted value. Although these data are meager they suggest that the apparent force of this type of muscular pull may follow a power function that is like the veg scale.

Auditory flutter refers to the apparent rate at which a white noise is turned on and off. Over a wide range of flutter frequencies (5 to 200 per second) the apparent rate of flutter grows as the 1.7 power of the frequency of stimulus interruption. Pollack used both fractionation and multiplication in determining this function. The one value he tested at a rate below 5 per second suggests that there may be a discontinuity in the function at about this point. Otherwise the data give a close fit to a power function.

Velocity of seen movement gave data that fit a power function so well that, as the authors say, "The goodness of fit may be judged as excellent." The subjects alternately viewed two horizontally moving targets consisting of vertical white and black bands, and they adjusted the velocity of one to make it appear half the other. The standard velocities ranged from 1.8 to 42 mm/sec.

Ekman and Dahlbäck (9) call attention to another interesting fact, namely, that when the subjects were asked to set one velocity *equal* to a standard velocity the variability of the settings was not very different from the variabilities obtained when one velocity was set to appear *half* of a standard. They say, "the degree of uncertainty in fractionation is only a little greater than that in equating stimuli." Hanes (28) had earlier noted the same thing in his experiments on brightness. He says, "Extremely interesting is the fact that the standard deviations around the group means of the equality matches . . . are of the same order of magnitude as the deviations around the means of the $\frac{1}{2}$ and $\frac{1}{3}$ estimates at the same brightness levels." I can add for loudness that the variability of halvings and doublings is actually slightly less, in general, than the variability of equality matches made between two noises of substantially different spectra, say, a hiss and a rumble. In experiments involving several thousand loudness matches (61) the standard deviations ranged from 1.5 db, for very similar spectra, to 9.0 db for disparate spectra. Comparable variabilities for 36 subjects who halved and doubled the loudness of a white noise averaged 4.2 db (45). For pure tones, Garner (14) reports variabilities for halving that range from about 3 to 6 db.

Visual flash rate gives the largest exponent yet encountered. For rates between 5 and 10 per second the exponent is about 2.0. Actually the subjective scale for flash rate seems to break up into two segments: one for slow rates where the subject probably judges time elapsed between flashes, and one for faster rates (greater than about 5 per second) where the subject seems to judge rapidity of flash. Except near the transition point, both segments of the scale approximate power functions with slopes decidedly greater than unity.

This completes the list. Let us now look briefly at some of the antecedents of our power law.

Some Antecedents

The idea that the relation between stimulus and perceptual response might be a power function is no new thing. What is new is the devising of experiments to demonstrate the fact.

The first recorded notion that a power function might be involved appears to be the conclusion reached by Plateau (41) in the 1850's—before the appearance of Fechner's *Elemente* (1860). Plateau reasoned that, since the apparent relations among different shades of gray remain *sensiblement le même* when the general illumination is changed, the *ratios* among the sensations produced by the grays must remain fixed. This, he said, and quite rightly, is more rational than Fechner's view (later revived by Wright [75]) that it is the *differences* that remain constant. Plateau's assumption entails a power function, Fechner's a logarithmic function. By the present evidence, Plateau was right and Fechner was wrong. Sad to relate, however, Plateau later changed his mind, and for a reason we now can see was irrelevant to the issue: Delboeuf's equisection experiment did not yield a power function. And Fechner, crafty polemicist that he was, disposed of Plateau's power function by the equally irrelevant argument that its champion had repented.

But the power function came back to plague Fechner by another route. Brentano (4) suggested that an increase in a sensation is just noticeable when the increase is a fixed proportion of the original sensation. Brentano did not explore the mathematical consequences of this introduction of a kind of "Weber's law" into the realm of sensation, but Fechner did the mathematics for him (10) and showed that, if Weber's law holds for *both* the sensations and the stimuli, the consequence is a power function. This way of getting to a power function was later argued for by others, particularly Grotenfelt (20), but since there is scarcely any way to prove the basic assumptions involved, the cause was lost under Fechner's massive attacks upon it.

A third approach to the power function was devised by Guilford (22, 23), who called it the "nth-power law." It is based on the Fechnerian premise that j.n.d.'s are subjectively equal and that the scale of sensation can be had by summing them up. Guilford's new twist was to alter Weber's law by making the stimulus increment ΔS proportional not to S but to a power of S. A Fechnerian integration then yields a power function—which demonstrates that we can sometimes reach a correct conclusion by starting from two wrong assumptions. The assumption that ΔI is a power function of I, with an exponent lying between 0.5 and 1.0, is not very easily justified by the facts. It is not so easily justified as the more plausible assumption that $\Delta S = k\ (S + d)$ where k and d are constants. This latter formula fits the data Guilford himself cited better than does the power function, and there is other evidence to support it (33).

But it is Guilford's other assumption that is the greater source of trouble. How serious is the consequence of assuming the equality of j.n.d.'s becomes clear if we begin with any of the power functions (Table 1) whose exponent is greater than 1.0 and follow Guilford's development in the reverse direction. We reach the startling conclusion "that ΔS would be a decreasing function of S, which is unheard of in psychophysical research" (24).

In the light of what we now know, it would appear that of the three historical approaches to the power function Plateau's was based on the most valid assumption. But unfortunately Plateau's premise was only an assumption, and its empirical verification had to wait for nearly a hundred years.

Our position now is more fortunate. The road to the formulation of a general psychophysical law need no longer be strewn with assumptions about j.n.d.'s or contrast steps. The principle I am proposing, that equal sensation ratios are produced by equal stimulus ratios, is merely the summary statement of what we observe on at least a dozen perceptual continua. The power function which this simple principle entails appears to be the first-order relation between stimulus and response. Second-order departures from this law are certain to exist, and their exploration may sometimes prove interesting and instructive, but the wide invariance of the first-order relation deserves to remain the matter of first importance.

THE SUBJECTIVE SIZE OF THE J.N.D.

The issue raised by Fechner when he assumed that each j.n.d. represents a constant increment in sensation is a stubborn and vexatious problem. It is equivalent to the assumption that measures of resolving power provide equal units that can be used as a measure of magnitude. Applied to Class I (prothetic) continua the assumption is dead wrong, but how can we prove it? Or rather, how can we make the proof persuasive and convincing? Proof has been offered in the past (33, 38, 52) but ingenious counter-arguments have been marshaled against the evidence. Piéron (40) is the objector most persistent and erudite, but others join the chorus. It seems so simple and sensible to take the j.n.d. as a unit of sensation. How can it not be proper?

It is improper simply because it is wrong. Let us consider a concrete case. Both fractionation and magnitude estimation show what is quite obvious to begin with, namely, that the apparent length of a line seen against a homogeneous background is very nearly a linear function of physical length. Two inches look about twice as long as one inch. So we know that subjective length is related to physical length by a power function whose exponent is close to unity. Now what about the j.n.d., the resolving power? We can easily tell visually when one centimeter is added to one centimeter, but we can scarcely tell when a centimeter is added to a meter. So resolving power must be relative, or approximately so. As Münsterberg (37) showed, Weber's law holds fairly well for length. Now we put these two facts together and what happens? We see immediately that in terms of their psychological magnitude, as measured by the scale we first set up by fractionation and magnitude estimation, the j.n.d.'s get larger as we go up the scale. Q.E.D., the j.n.d. is not constant in subjective size, and counting them off does not yield the magnitude scale.

The principle involved can be clearly seen if we put the problem in more general terms. Let us assume that sensation ψ is a power function of stimulus magnitude S and also that Weber's law is true, i.e., $\Delta S = kS$. By the first assumption we may write, disregarding constants that depend only on the choice of units,

$$\psi = S^n.$$

If Weber's law is true, when we count off the j.n.d.'s and record their number J as a function of the stimulus, we obtain

$$J = \log_c S,$$

where c is a constant that can serve as the base of the logarithm. We can rewrite this equation as

$$S = c^J$$

and substitute this expression for S in the first equation, which gives

$$\psi = c^{Jn}.$$

This equation tells us that the psychological magnitude grows as an exponential function of the number of j.n.d.'s J, which brings us to the interesting conclusion that the scale of sensation is an exponential function of the scale Fechner believed it to be. If we differentiate this equation we get

$$\frac{d\psi}{dJ} = ac^{Jn},$$

where a is a constant. The equation tells us that instead of being constant as Fechner assumed, the subjective size of the j.n.d. grows as an exponential function of the number of j.n.d.'s above threshold. Simple substitution in this last equation also reveals that the subjective size of the j.n.d. is proportional to the psychological magnitude ψ.

These equations tell us how it would be if Weber's law were strictly true, but since this law is not strictly true the situation is not quite so simple. The complications, however, are only second order. A superior form of Weber's law is $\Delta S = k(S + d)$, which was known to Fechner, Helmholtz and others. By the addition of the small constant d, this form makes allowance for the inevitable residual "noise level" in the perceptual system, but at the same time it preserves the basic principle that resolving power is relative. If this form is accepted, the foregoing equations and derivations get complicated, but the conclusions they lead to are not substantially altered. It is still true that the subjective size of the j.n.d. grows rapidly as we go up the scale.

Actually, since the value of the constant d is relatively small, its addition to the formula for Weber's law affects the functions in a significant way only at low stimulus levels. At levels near threshold the general formulas developed above break down, but over the middle and higher stimulus ranges they give a good account of things.

A concrete instance of these general relations is shown graphically in Fig. 7. Miller (33) measured the j.n.d. for white noise by the so-called quantal procedure (65) which provides an estimate of the quantal increment in the stimulus that is just detectable. The results are well described by the relation $\Delta I = k(I + d)$. Miller summated the quanta (increments) and obtained the top curve shown in Fig. 7. Except at the low end it is almost a straight line when plotted against decibels. The bottom curve is the sone scale of subjective loudness. The triangles represent the median results obtained by J. C. Stevens and E. Tulving (51) from 70 subjects who made direct magnitude estimations

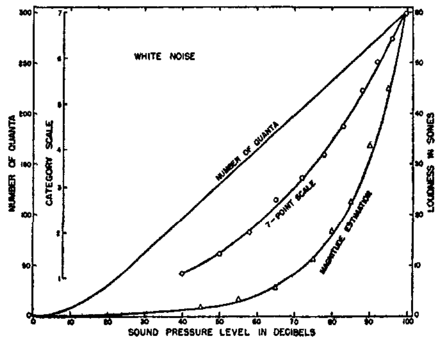


FIG. 7. J.n.d. (quantum) scale, category scale, and ratio scale (sones) for the loudness of white noise. *Circles:* with the spacing of the levels adjusted to give a "pure" category scale, 8 subjects each made two judgments of each level on a 7-point scale. The range was disclosed at the outset and the order of presentation was different for each subject. *Triangles:* median judgments by 70 subjects who made magnitude estimations of eight levels presented in irregular order in a classroom, with no designated standard. Estimates were transformed to a common modulus at the 75-db level, which was the first level presented.

of the loudness of different levels of white noise presented in irregular order. In a log-log plot these points fall close to a straight line like that shown in Fig. 6.

The curves in Fig. 7 show how vastly different are the two scales, the j.n.d. (or quantum) scale and the loudness scale. In this respect they confirm what we have already seen in Fig. 1. They also show that when we use the loudness scale to measure the subjective size of successive j.n.d.'s, their size grows rapidly as we go up the scale. Except near the low end, their growth approximates an exponential function. Actually, over the low end of the scale where the j.n.d. scale is curved there is a superficial similarity between the j.n.d. scale and the loudness scale, but over the major portion of the audible range the divergence between the two scales is large and obvious.

The middle curve in Fig. 7 shows an

example of the function obtained when the loudness of a white noise is judged on a scale from 1 to 7 (data from Stevens and Galanter [64], Fig. 9A). The spacing of the stimuli had been adjusted to give a "pure" category scale. Just as in Fig. 1, we see that the form of the category scale is intermediate between that of the j.n.d. scale and the scale of subjective magnitude. If plotted against the sone scale the category scale would be concave downward, as are all category scales on prothetic continua.

It has occasionally been asserted that some scales of sensory magnitude agree with the scale obtained by counting off j.n.d.'s (74). Hanes' (27) statement that this is true for brightness is a recent example, which has led, incidentally, to a certain amount of confusion. In a subsequent study Hanes (28) extended the range of his measurements far enough for us to see clearly that the scale of subjective brightness (in brils) departs from the j.n.d. scale in the expected way. Perhaps Hanes should have stated this explicitly.

I have summed the j.n.d.'s for brightness obtained by three different experimenters (2, 18, 35) and have compared the results with Hanes' bril scale. The pictures presented are essentially similar to that shown for loudness in Fig. 7, except that the curved tail at the lower end of the j.n.d. scale tends to be longer. The basic fact is that over the middle and upper ranges of intensity Weber's law holds fairly well for brightness, but the bril scale is a power function. Over the range of the fainter luminances, Weber's law breaks down to a degree sufficient to produce the apparent agreement Hanes spoke of. But it is quite wrong to jump to the general conclusion that j.n.d.'s add up to the bril scale.

We have already noted, in the discussion of Guilford's "nth-power law," that in order for the j.n.d.'s to add up to give the magnitude function, ΔS cannot in general be an increasing function of S. For most perceptual continua it must be a decreasing function of S. But this is "unheard of in psychophysical research." It would mean, for example, that it would be easier to tell when a centimeter was added to a meter than when a centimeter was added to a centimeter. Resolving power simply does not work that way.

Suppose now for a moment we risk Titchener's displeasure and try to think like a physicist. Physics also faces the problem of resolving power and in most physical measurements resolving power is relative—the accuracy of measurement is proportional to magnitude. This fact is expressed by saying that length can be measured to 1 part in 10^5, that resistance can be measured to 1 part in 10^6, or that time intervals can be measured to 1 part in 10^8. In this sense Weber's law has its analogue in physics. Now, it is conceivable that someone might propose that we give up our ordinary ways of measuring length, say, and devise instead a scale whose unit is the resolving power of our best procedures. We would then be doing what Fechner proposed. In terms of our present units, the new units of length would grow increasingly larger as we go up the scale, and they would do so for the same reason that the j.n.d.'s grow larger as we go up the psychological scale on a prothetic continuum. The laws of physics could be rewritten in terms of the new scales of units based on resolving power, but it would probably strike the physicist as less than sensible to try to do it—and not simply because it would be hard work. For reasons that he could readily justify, the physicist wants to start with magnitudes and not with resolving power. What I am suggesting is that there is sufficient parallel between physics and psychophysics to justify the same approach in both fields. In neither discipline should we try to make resolving powers, variabilities, confusions, j.n.d.'s, discriminal dispersions or any such entities serve as units of magnitude on continua that behave like the prothetic continua of Class I.

SCALING THEORY

The same counsel recommends itself to those of us who devise scaling procedures for other than perceptual variables. As Gulliksen (26) complains, those who worry about psychophysical measurement often ignore the work on nonsensory scaling, and vice versa. Titchener (71) also once complained that psychologists have ignored the advances in subjective scaling achieved by the meteorologists. Insularity is probably inevitable in an age of specialization, but it has its dangers. One of the dangers is that the laws of human judgment developed under simple conditions, where we can better see what is going on, may be ignored when we work in more complicated realms where discernment is more difficult.

A large segment of psychological scaling theory has been built around models that make scale "units" out of some measure of "discriminal dispersion," but this is precisely the kind of unit that is demonstrably *not* invariant on prothetic continua. Variabilities and confusions provide a tempting starting point for psychological measurement, chiefly because dispersion among our judgments is something we always have with us, but the transforming of unreliability, inconstancy, or confusion into units of measure somehow needs more than assumptions to back it up. Our fundamental scales can, I think, stand on firmer ground than the mere scatter of our data.

It is recognized, of course, that measures of confusion often have a usefulness of their own, and that for certain purposes it is desirable to construct "equal discriminability" scales (16). The engineering of schemes for optimizing the coding of information on cathode-ray tubes presents an interesting example of how such measures may be used in the prediction and control of confusability (36). But our concern here is with a different problem, namely, whether measures of confusion can be used as units of magnitude.

The use of dispersion as a tool for magnitude scaling requires assumptions regarding its behavior. How does dispersion (in psychological units) vary over the psychological continuum in question? A variety of assumptions regarding the behavior of variability have been elaborated, but so far as I know, none of them come close to admitting the drastic growth in dispersion that we find to be true for Class I (prothetic) continua. Here, we recall, the dispersion (in subjective units) approaches an exponential function of the number of j.n.d.'s above threshold or, alternatively stated, it grows almost in direct proportion to the psychological magnitude. The assumption made in practice is usually that dispersion (in psychological units) is constant, or nearly so. This is like Fechner's assumption that j.n.d.'s are equal in subjective size. Direct experiments can sometimes verify this assumption on continua of Class II (metathetic), but on the "quantitative" continua of Class I they show that the assumption fails by a wide margin.

On continua that behave like Class I we would be closer to right if we began with the assumption that discriminal dispersion is not constant but is proportional to the psychological magnitude in question. When the psychological magnitude is a power function of the stimulus, this assumption is equivalent to saying that psychological values separated by equal units of dispersion on the stimulus scale stand in a constant ratio to each other. If we were to start from this assumption, could we then proceed to construct a useful scale of the psychological magnitude from measurements made of discriminal dispersion? In principle we might erect a new type of scale, but we could not construct an interval or a ratio scale of the kinds we ordinarily employ. While working on the theory of scales (54) back in the late 1930's I had a hard time convincing myself that if we know concerning a series of values only that $a/b = b/c = c/d = \ldots$ we are powerless to proceed further toward what I have called a ratio scale of measurement. But the fact is that the equating of ratios gets us no further unless (a) we can also equate intervals, or (b) we can determine the numerical value of the ratio (as we do under the ratio methods reviewed above). An example of how we might use equated ratios *plus* equated intervals to construct a ratio scale is discussed elsewhere (55, p. 24).

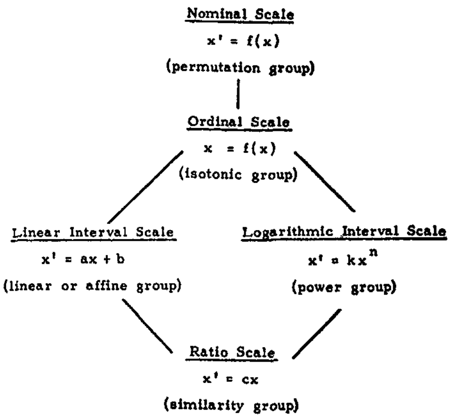
But let us pursue for a moment the problem of what we might do when we have equated a set of ratios: $a/b = b/c = c/d \ldots$ We can assume that we have an operation for ordering these values and that $a < b < c < d \ldots$ The problem then is, how may we assign numerical values to this series? As is true when we have a series of equated intervals, $a - b = b - c = c - d = \ldots$, we can pick any two values arbitrarily (subject only to the ordinal requirement), and having done so, all the other values are determined. Thus for the ratio series, if $a = 2$ and $b = 6$ then $c = 18$, $d = 54$, etc. The next question is, what transformations are permitted on this scale? We can show that any value x may be replaced by x' provided $x' = kx^n$ where k and n are constants that may take on any positive values. In other words we can always transform the scale values by a power function. For example, if we square all the values, the ratios remain equal.

A scale of this kind may be mathematically interesting, but, like many mathematical models, it has thus far proved empirically useless. At least, so far as I am aware, no uses for it have yet been found. In a formal sense there are certain parallels between a scale of equal but indeterminate ratios and a scale of equal but indeterminate intervals. On both, for example, any two values can be selected arbitrarily, and the scale values of both can be transformed by an equation involving two arbitrary constants. But the scale of equal intervals has great utility, as witness the Fahrenheit and Celsius scales of temperature, and the scale of calendar time. And differences on interval scales can be measured on ratio scales, as when we measure intervals of duration in seconds. But what can we do with equated ratios when the value of the ratio is indeterminate?

An obvious suggestion is that we might convert to logarithms and express the equated ratios as $\log a - \log b = \log b - \log c = \log c - \log d$, etc. If we then restrict the values of $a, b, c, \ldots$ to positive numbers, we can set up an interval scale in logarithms. There is no a priori reason why we could not put this scale to work in a manner analogous to the workings of our linear interval scales. The linear and the logarithmic scales have in common the fact that the choice of the zero point is arbitrary. It might be interesting to explore the consequences of this procedure and perhaps try to answer the question why science has never made use of these logarithmic interval scales based on equated ratios.

Another question arises. If we were to develop such scales, what would we call them? When I originally proposed (54) the names *nominal, ordinal, interval,* and *ratio* for the four classes of scales commonly used, I worried some about the scale based on equated ratios, but I dismissed it as a purely academic question, for there appeared to be no examples of its use. If uses for it were to develop, we would like, perhaps, to call this class of scales *ratio* scales, for they are based on an empirical operation for equating ratios. But since the term ratio scale has now gained wide currency as the name for the scales that admit only a similarity transformation (multiplication by a constant) it may be better to call the scale based on equated ratios a *logarithmic interval* scale.

With this class of scales included, the hierarchy of scales, together with the group of transformations permitted by each scale, would take the form shown in the diagram.



As this diagram suggests, a ratio scale is possible whenever empirical operations are available to create both types of interval scales. If we can determine both equal differences and equal ratios we can eliminate the additive constant b and the exponent n, and we are left with only a multiplying constant c. This defines a scale with a zero point—what we call a ratio scale, or, if we prefer, a "zero-point" scale.

Returning now to the assumptions made about discriminal dispersion we note that, in terms of the scales listed in the diagram, the assumption most often made is that the processing of variability will yield a linear rather than a logarithmic interval scale. It is not only in connection with the method of pair comparisons, which Thurstone called the "best of all the psychophysical methods" (70), that the wrong assumption of constant dispersion (in psychological units) often gets made. This assumption is either latent or explicit in many subtler procedures, such as the methods of "successive intervals," "graded dichotomies," "successive categories," etc. (see 23). The foregoing arguments suggest that whenever these methods are applied to prothetic continua, continua on which the j.n.d. (dispersion) tends to vary directly with magnitude, the resulting scales will turn out to be nonlinear. It seems obvious that none of the scales listed in Table 1 could be constructed by the method of pair comparisons, or by any of its several cousins. The laws of judgment derived from direct experiments on perceptual continua make it plain that, unless it can be demonstrated that the judgmental continuum behaves like those of Class II (metathetic), any scaling procedure that is geared to the assumption that equally often noticed differences are equal is defective.

Fortunately, this does not mean that no scaling is possible. We still have other devices, such as the direct methods for ratio scaling. Is there a genuine, nontrivial, substantive problem calling for a psychological scale for which these methods, or some variation on them, cannot produce an acceptable answer? Or is it merely that we like to assume, with Fechner, that indirection is the best path to the goal? Contrary to a common assumption, the use of the direct methods does not require knowledge of an underlying, measurable physical continuum (48). Only a nominal scale (55) is required at the stimulus level, i.e., the stimuli must be identifiable by the experimenter. For example, when the observer judges the apparent ratio between two stimuli the experimenter needs to know which two stimuli were involved, but he does not need to know their values on any other scale. I know from experience that it is more comfortable to take it for granted that a direct method is impossible than it is to try to work one out. And it is still more onerous to test and refine the direct methods and to purge them of constraints and biases. The problem of the scaling of psychological continua is full of booby traps, but so is the laboratory measurement of a physical continuum like, for instance, electrical inductance or mechanical impedance.

Perhaps one of our professional idiosyncracies is that in psychology we are sometimes more enamored of models and methods than we are of problems. One occasionally gets the impression that there are more people with a method who are looking for a problem to use it on than there are searchers with a problem looking for a method. It is admittedly more entertaining to make a formal model, complete with assumptions, postulates, and derivations, than it is to grub through the empirical tangle of an experimental issue. But if something empirically useful is to issue from a model, something empirically known must be put into it. The gratuitous assumption that some measure of dispersion, sensitivity, or resolving power can be used as the unit of a scale of psychological magnitude does not meet this requirement.

SUMMARY

This paper has rambled far enough afield at this point and it is time to try to restate the issues in briefer compass. The main points are these.

Two general classes of perceptual continua can be distinguished by means of four functional criteria. On Class I or "quantitative" continua the j.n.d. increases in subjective size as psychological magnitude increases, category rating-scales are concave downward when plotted against psychological magnitude, comparative judgments exhibit a time-order error (a "category effect"), and equisection experiments exhibit hysteresis. On Class II or "qualitative" continua these four effects are apparently absent. Class I, called prothetic, includes those continua on which discrimination is mediated by an additive mechanism at the physiological level; Class II, called metathetic, includes those mediated by a substitutive mechanism.

On Class I (prothetic) continua the use of one or more of four kinds of direct methods for constructing ratio scales reveals that equal stimulus ratios tend to produce equal subjective ratios. Hence, to a first-order approximation the "psychophysical law" relating stimulus and response is a power function. The exponent, as measured on fourteen different continua, varies from about 0.3 for loudness to about 2.0 for visual flash rate. A few workers in the past have conjectured this power function, even starting sometimes from wrong assumptions. Only lately, however, is this law becoming securely anchored in experiment.

Fechner's logarithmic law is not found in experiment for the simple reason that resolving power (the j.n.d.) is not constant in psychological units, but is roughly proportional to psychological magnitude. For this reason, all procedures of Fechnerian extraction, like the method of pair comparisons and its related techniques, which seek to build scales out of "unitized" measures of dispersion, are not proper methods for scaling magnitudes that behave like Class I or prothetic continua. Whenever psychological scales are called for, direct ratio scaling methods should probably be tried.

REFERENCES

1. BEEBE-CENTER, J. G., & WADDELL, D. A general psychological scale of taste. *J. Psychol.*, 1948, **26**, 517–524.
2. BLACKWELL, H. R. Contrast thresholds of the human eye. *J. opt. Soc. Amer.*, 1946, **36**, 624–643.
3. BORING, E. G. The stimulus-error. *Amer. J. Psychol.*, 1921, **32**, 449–471.
4. BRENTANO, F. *Psychologie vom empirischen Standpunkte.* Th. 1. Leipzig: Dunker & Hunblot, 1874.
5. COHEN, J., HANSEL, C. E. M., & SYLVESTER, J. D. Interdependence of temporal and auditory judgments. *Nature,* 1954, **174**, 642.
6. CORBIN, H. H., REESE, E. P., REESE, T. W., & VOLKMANN, J. Experiments on visual discrimination, 1952–55. Psychophysical Research Unit, Mt. Holyoke College, AFCRC TR No. 56-52, April 1956.
7. COWDRICK, M. The Weber-Fechner law and Sanford's weight experiment. *Amer. J. Psychol.,* 1917, **28**, 585–588.
8. EKMAN, G., BERGSTROM, B., & KÜNNAPAS, T. M. Comparison between two psychophysical scaling methods. Report from the Psychological Laboratory, Univer. of Stockholm, No. 37, August 1956.
9. EKMAN, G., & DAHLBÄCK, B. A subjective scale of velocity. Report from the Psychological Laboratory, Univer. of Stockholm, No. 31, February 1956.
10. FECHNER, G. T. *In Sachen der Psychophysik.* Leipzig, 1877.
11. FERNBERGER, S. W. On absolute and relative judgments in lifted weight experiments. *Amer. J. Psychol.,* 1931, **43**, 560–578.
12. Final Report. *Advancement of Science,* 1940, No. 2, 331–349.
13. FULLERTON, G. S., & CATTELL, J. McK. *On the perception of small differences.* Philadelphia: Univer. of Pennsylvania Press, 1892.
14. GARNER, W. R. Some statistical aspects of half-loudness judgments. *J. acoust. Soc. Amer.,* 1952, **24**, 153–157.
15. GARNER, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.,* 1954, **26**, 73–88.
16. GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.,* 1951, **58**, 446–459.
17. GILINSKY, ALBERTA S. Perceived size and distance in visual space. *Psychol. Rev.,* 1951, **58**, 460–482.
18. GRAHAM, C. H., & BARTLETT, N. R. The relation of size of stimulus and intensity in the human eye. III. The influence of area on foveal intensity discrimination. *J. exp. Psychol.,* 1940, **27**, 149–159.
19. GREGG, L. W. Fractionation of temporal intervals. *J. exp. Psychol.,* 1951, **42**, 307–312.
20. GROTENFELT, A. *Das Webersche Gesetz und die psychische Relativität.* Helsingfors: Frenckell, 1888.
21. GRUBER, H. E. The relation between perceived size and perceived distance. *Amer. J. Psychol.,* 1954, **67**, 411–426.
22. GUILFORD, J. P. A generalized psychophysical law. *Psychol. Rev.,* 1932, **39**, 73–85.
23. GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.
24. GUILFORD, J. P., & DINGMAN, H. F. A validation study of ratio-judgment methods. *Amer. J. Psychol.,* 1954, **67**, 395–410.
25. GUILFORD, J. P., & DINGMAN, H. F. A modification of the method of equal-appearing intervals. *Amer. J. Psychol.,* 1955, **68**, 450–454.
26. GULLIKSEN, H. Paired comparisons and the logic of measurement. *Psychol. Rev.,* 1946, **53**, 199–213.
27. HANES, R. M. A subjective brightness scale. *J. exp. Psychol.,* 1949, **39**, 438–452.
28. HANES, R. M. The construction of subjective brightness scales from fractionation data: a validation. *J. exp. Psychol.,* 1949, **39**, 719–728.
29. KARLIN, J. E., & STEVENS, S. S. Auditory tests for the ability to discriminate the pitch and loudness of noises. Report from the Psycho-Acoustic Laboratory, Harvard Univer., OSRD No. 5294 (PB 39576), 1 August 1945.
30. MACLEOD, S. A construction and attempted validation of sensory sweetness scales. *J. exp. Psychol.,* 1952, **44**, 316–323.
31. MERKEL, J. Die Abhängigkeit zwischen Reiz und Empfindung. *Phil. Stud.,* 1888, **4**, 541–594; 1889, **5**, 245–291; 1889, **5**, 499–557.
32. METFESSEL, M. F. A proposal for quan-

titative reporting of comparative judgments. *J. Psychol.,* 1947, **24**, 229–235.

33. MILLER, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.,* 1947, **19**, 609–619.
34. MOON, P. *The scientific bases of illuminating engineering.* New York: McGraw-Hill, 1936.
35. MUELLER, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. *J. gen. Physiol.,* 1951, **34**, 463–474.
36. MULLER, P. F., *et al.* Symbolic coding of information on cathode ray tubes and similar displays. *WADC Tech. Rep. 55–375.* Wright Air Development Center, Ohio, 1955.
37. MÜNSTERBERG, H. A psychometric investigation of the psychophysic law. *Psychol. Rev.,* 1894, **1**, 45–51.
38. NEWMAN, E. B. The validity of the just noticeable difference as a unit of psychological magnitude. *Trans. Kansas Acad. Sci.,* 1933, **36**, 172–175.
39. NEWMAN, E. B., STEVENS, S. S., & VOLKMANN, J. On the method of bisection and its relation to a loudness scale. *Amer. J. Psychol.,* 1937, **49**, 134–137.
40. PIÉRON, H. *The sensations, their functions, processes and mechanisms.* New Haven: Yale Univer. Press, 1952.
41. PLATEAU, M. H. Sur la mesure des sensations physique, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bull. de l'Acad. Roy. Belg.,* 1872, **33**, 376–388.
42. POLLACK, I. Auditory flutter. *Amer. J. Psychol.,* 1952, **65**, 544–554.
43. POSTMAN, L. The time-error in auditory perception. *Amer. J. Psychol.,* 1946, **59**, 193–219.
44. POSTMAN, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.,* 1947, **60**, 101–108.
45. POULTON, E. C., & STEVENS, S. S. On the halving and doubling of the loudness of white noise. *J. acoust. Soc. Amer.,* 1955, **27**, 329–331.
46. PRATT, C. C. Time-error in the method of single stimuli. *J. exp. Psychol.,* 1933, **16**, 798–814.
47. REESE, E. P. (Ed.) Psychophysical research summary report, 1946–1952. Psychophysical Research Unit, Mt. Holyoke College. USN Spec. Dev. Ctr. *Tech. Rep. No. SDC-131-1-5,* 1953.
48. REESE, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.,* 1943, **55**, No. 3 (Whole No. 251).
49. RICHARDSON, L. F., & ROSS, J. S. Loudness and telephone current. *J. gen. Psychol.,* 1930, **3**, 288–306.
50. ROSS, S., & KATCHMAR, L. The construction of a magnitude function for short time intervals. *Amer. J. Psychol.,* 1951, **64**, 397–401.
51. STEVENS, J. C., & TULVING, E. Loudness estimation in a group situation. *Amer. J. Psychol.* (in press).
52. STEVENS, S. S. A scale for the measurement of a psychological magnitude loudness. *Psychol. Rev.,* 1936, **43**, 405–416.
53. STEVENS, S. S. On the problem of scales for the measurement of psychological magnitudes. *J. Unif. Sci.,* 1939, **9**, 94–99.
54. STEVENS, S. S. On the theory of scales of measurement. *Science,* 1946, **103**, 677–680.
55. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
56. STEVENS, S. S. Pitch discrimination, mels, and Kock's contention. *J. acoust. Soc. Amer.,* 1954, **26**, 1075–1077.
57. STEVENS, S. S. On the averaging of data. *Science,* 1955, **121**, 113–116.
58. STEVENS, S. S. The measurement of loudness. *J. acoust. Soc. Amer.,* 1955, **27**, 815–829.
59. STEVENS, S. S. Decibels of light and sound. *Physics Today,* 1955, **8** (10), 12–17.
60. STEVENS, S. S. The direct estimation of sensory magnitudes: loudness. *Amer. J. Psychol.,* 1956, **69**, 1–25.
61. STEVENS, S. S. The calculation of the loudness of complex noise. *J. acoust. Soc. Amer.,* 1956, **28**, 807–832.
62. STEVENS, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.* (in press).
63. STEVENS, S. S., & DAVIS, H. *Hearing: its psychology and physiology.* New York: Wiley, 1938.
64. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. Report from the Psycho-Acoustic Laboratory, Harvard Univer., PNR–186, August 15, 1956. (Submitted to *J. exp. Psychol.*)
65. STEVENS, S. S., MORGAN, C. T., & VOLK-

MANN, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.,* 1941, **54**, 315–355.

66. STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.,* 1956, **51**, 71–78.
67. STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.,* 1940, **53**, 329–353.
68. STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.,* 1937, **8**, 185–190.
69. TAVES, E. H. Two mechanisms for the perception of visual numerousness. *Arch. Psychol., N. Y.,* 1941, No. 265.
70. THURSTONE, L. L. Fechner's law and the method of equal appearing intervals. *J. exp. Psychol.,* 1929, **12**, 214–224.
71. TITCHENER, E. B. The psychophysics of climate. *Amer. J. Psychol.,* 1909, **20**, 1–14.
72. TITCHENER, E. B. *Experimental psychology.* Vol. II, Part II (Instructor's Manual). New York: Macmillan, 1923.
73. WOODROW, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Chap. 32.
74. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.
75. WRIGHT, W. D. *Researches on normal and defective colour vision.* St. Louis: Mosby, 1947.

(Received for early publication November 2, 1956)