Psychological Review
1971, Vol. 78, No. 5, 426–450

ISSUES IN PSYCHOPHYSICAL MEASUREMENT[1]

S. S. STEVENS[2]

Harvard University

Two classes of ratio-scaling procedures are outlined—magnitude matching and ratio matching—and their assets and liabilities are noted. Partition-scaling procedures, which are supposedly designed for interval scaling, produce results that can be described by a power function with a virtual or "as if" exponent. Since the virtual exponent is smaller than the actual exponent of the continuum, the category scale is nonlinear. The virtual exponent provides a convenient descriptor of several kinds of partition operations. Other topics discussed include individual differences among subjects' exponents, procedures of averaging, and the effects of stimulus range on exponents. It is suggested that the power law asserts a nomothetic imperative.

The task here is to review the matching procedures used to determine the power functions that govern the growth of sensation magnitude and to consider some of the sources of deviation and perturbation that have raised questions concerning the nomothetic quality of the psychophysical power law.

Since all procedures of measurement involve matching operations, the interesting differences among different scales and different kinds of measurement can often be reduced to a basic question: What was matched to what, and how? In the domain of psychophysics, numerous scaling methods have been invented, many of them useful for the determination of ratio scales of apparent magnitude. The approaches of ratio scaling can be catalogued in different ways, but for present purposes they fall into two general classes: *magnitude matching*, which includes the subclasses (*a*) cross-modality matching, (*b*) magnitude estimation, and (*c*) magnitude production; and *ratio matching* which includes the subclasses (*a*) cross-modal ratio matching, (*b*) ratio estimation, and (*c*) ratio production.

Since there are endless variations on psychophysical procedures, it is possible here to comment on only a few of the features that characterize the principal methods. Consideration is also given to some of the interval or partition methods, and to the distinction between the virtual exponent and the actual exponent. Other problems discussed concern individual differences, averaging, and range effects.

[1] Supported in part by Grant NS-02974 from the National Institutes of Health (Laboratory of Psychophysics Report ppr-366-133).

[2] Requests for reprints should be addressed to S. S. Stevens, Laboratory of Psychophysics, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Magnitude Matching

These procedures include all direct equations between two continua. Three principal varieties of magnitude matching have been distinguished.

Cross-Modality Matching

When the stimulus for a continuum can be readily varied by means of a control of some kind, it becomes possible to match that continuum to any other continuum. Figure 1 gives examples of matching functions produced when several different continua were matched to vibration on the fingertip.

Ideally, the experiment comparing two continua should provide for a balanced design in which each continuum is matched in turn to the other continuum. A balanced procedure may help to assess and correct the regression effects that are always present in the matching operation (Stevens & Greenbaum, 1966). In the typical experiment, the observer tends to shorten the range of whichever variable he controls. Even within the same modality the regression effect shows up in matching functions. Thus, two somewhat

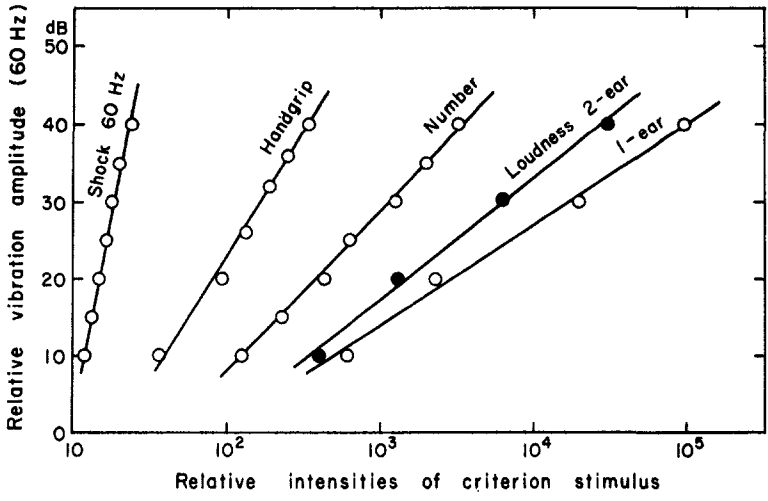

FIG. 1. Equal sensation functions obtained by cross-modality matches between various continua and a 60-Hz. vibration on the fingertip. (The vibration amplitudes were set by the experimenter. The observer adjusted the other stimulus to produce an apparent match—from Stevens, 1968a.)

different functions were obtained, depending on which auditory stimulus the subject adjusted in matching a tone to a noise.

Because it is often difficult to give the subject control of the stimulus, many cross-modality comparisons have not yet been made. It is difficult, for example, for the subject to vary the heaviness of lifted weights in order to match heaviness to loudness.

When a balanced design in the matching of two continua is impracticable, an evaluation of the regression effect may sometimes prove possible by way of a third continuum. Two principal paradigms can be distinguished.

1. Continuum A is adjusted to match each of two continua, B and C. The ratio of the exponents of the matching function A to B and A to C determines the exponent of the function relating B to C. The derived function would presumably be free of the regression effect, provided the regression occasioned by adjusting A remained constant when the criterion continuum was changed from B to C. The hoped-for constancy might be upset by such factors as disparities in difficulty or range.

2. A second procedure for counterbalancing the regression effect can be utilized whenever it is possible to match both A and B to a common continuum C. The two matching functions provide exponents whose ratio determines the exponent of the power function relating A to B. Whether the regression effects in the two matching functions A to C and B to C are exactly equal may not be known, of course, but the procedure may still cancel a major portion of the regression bias.

An instance of the second paradigm was provided by Moskowitz (1969), who asked observers to match both numbers and loudness to a wide variety of taste mixtures. From each of the 68 pairs of experiments that Moskowitz conducted we can derive an estimate of the power-function exponent relating number to loudness. The geometric mean of the 68 estimates was .67, which agrees with the value of the exponent adopted for loudness calculation (Stevens, in press). The standard deviation was .51 decilog, or about 12%. That degree of scatter may be regarded as an empirical guide to the amount of variability to be expected in tests of transitivity among the exponents of the power functions. When the regression effect has been canceled, about two-thirds of the measured exponents may be expected to lie within plus or minus half a decilog of the predicted exponents (see Stevens, 1969).

The regression effect, of course, is only one of the sources contributing to the systematic errors that affect the outcome of experiments, but it is one of the most obstinate, and therefore perhaps the most important. And it may be composed of more than a single factor.

Magnitude Estimation

This procedure is actually a form of cross-modality matching in which numbers are matched to stimuli. When first used, the procedure was called absolute judgment (Stevens, 1953), later numerical estimation (Stevens, 1954), and still later magnitude estimation (Stevens, 1955a). That last name appears to have stuck. In this context, the number continuum can be regarded as another perceptual modality. Magnitude estimation or "number matching" has become a popular method, mainly because of its convenience. The subject brings the numbers with him, so to speak, and the experimenter needs only to provide the target stimuli to which the numbers are to be matched. The nature of the task can be portrayed in terms of a typical set of written instructions to the subject.

> You will be presented with a series of stimuli in irregular order. Your task is to tell how — they seem by assigning numbers to them. Call the first stimulus any number that seems to you appropriate. Then assign successive numbers in such a way that they reflect your subjective impression. For example, if a stimulus seems 20 times as —, assign a number 20 times as large as the first. If it seems one-fifth as —, assign a number one-fifth as large, and so forth. Use fractions, whole numbers, or decimals, but make each assignment proportional to the — as you perceive it.

Experience has shown that it is usually better not to designate a standard. The subject then remains free to choose his own modulus. If possible, stimuli should be presented in a different irregular order to each subject, but the first stimulus is usually chosen from among those in the middle region, rather than from one end of the range. Between 10 and 20 stimuli may be presented at a session. A good schedule provides for one judgment, or at the most two judgments, per stimulus by each subject. After the subject has learned to recognize a particular stimulus, little or no new information is obtained from subsequent judgments of its repeated presentation. Furthermore, biases due to range and spacing of stimuli seem to have less effect when the subject is limited to one judgment per stimulus.

Untrained, inexperienced college subjects seem to do as well at the matching tasks as those who have had many years of practice. Hence, there is no need to "train" the subjects. Indeed, since there is no right or wrong to the subjects' responses, it is not clear what would be meant by training. Under some circumstances, the nature of the task may profitably be clarified by allowing the subjects to begin by matching numbers to an easier continuum, such as apparent length of lines, or apparent size of circles.

Averaging can be done by computing geometric means or medians. The log-log slope (exponent) determined by the geometric means is not affected by the fact that each observer uses a different unit of modulus. When it is desired to adjust the judgments to a common modulus, a good method is to minimize the squares of the individual subject's intercept differences. The procedure is: convert all scores to logs, compute grand mean of logs, and adjust each log score for each observer by whatever additive constant makes the observer's mean correspond to the grand mean. That procedure of *modulus equalization* permits each of an observer's estimates to contribute to the correction factor to be applied to that observer's modulus.

Magnitude Production

Here the experimenter presents the numbers one at a time in irregular order, and the subject adjusts the stimulus to produce an apparent match. The numbers themselves should normally approximate a geometrical progression. For example, in an extensive study of loudness and its inverse, softness, the successive numbers presented were in the ratio 2 to 1 and ranged from 1.25 to 80 (Stevens & Guirao, 1962). Sample results are shown by the triangles in Figure 2.

Because of the regression effect, the power functions obtained by magnitude production

are typically steeper (have larger exponents) than those obtained by magnitude estimation. An unusually large regression angle is illustrated in Figure 2. It is often assumed that the unbiased function lies between the two functions obtained by estimation and production, and indeed it may. But where? Are the error sources in the one procedure exactly balanced by the error sources in the other? Although exact balance may be possible, it seems hardly likely that no asymmetry exists. An average function may nevertheless be desired, in which case it may be well to compute the geometric mean of the two exponents. The geometric mean is invariant under an interchange of the two coordinates (Indow & Stevens, 1966). When the results of magnitude estimation and production are combined in an appropriate way, the combined procedure offers advantages over either procedure alone.

A particular version of the combined procedure designed to produce a balanced function has been spelled out by Hellman and Zwislocki (1968) and applied to loudness functions.

Another way of combining some of the features of production and estimation is to permit the subject to set stimulus levels at his own pleasure and to report the apparent magnitude. The experimenter in effect gives up all control over the stimuli. A radical procedure of that kind was used by J. C. Stevens and Guirao (1964) to show that individual subjects produce power functions and that the power function is not, as some writers had suggested, an artifact of averaging. The key parts of the instructions were as follows:

> Your task is to set the tone to different levels of loudness and to assign numbers to each of the loudnesses. Make your numbers proportional to the loudness you hear. You may make as many settings as you want. Try to cover a wide range of loudness.

Eleven subjects, chosen at random from among students, staff, and secretaries, gave the results shown in Figure 3. One subject made as few as seven production estimations. Another made five times that many. Two subjects made settings that extended over almost 100 decibels (db.) and made estimations that ranged over about 10,000 to 1, which implies an exponent near .8. The experiment was repeated some months later. The geometric mean of the 22 exponents for the 11 subjects was .7, which is fairly close to the value 2/3, which has been proposed as the standard value (Stevens, in press).

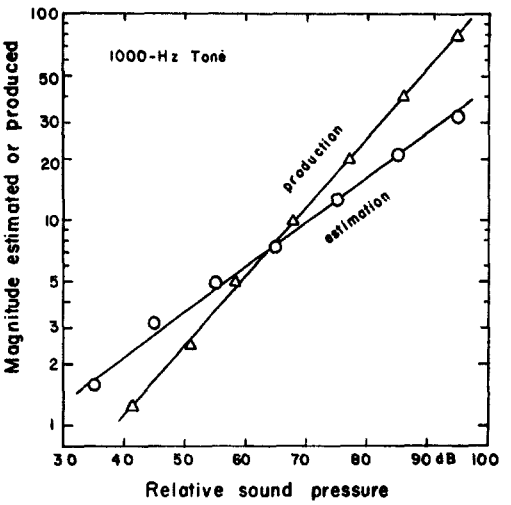

FIG. 2. Magnitude estimation and magnitude production of loudness. (Each point is the geometric mean of two estimates or two productions by each of 10 observers.) (Reprinted with permission from an article by J. C. Stevens and M. Guirao published in the *Journal of the Acoustical Society of America*, 1964, Vol. 36. Copyrighted by the Acoustical Society of America, 1964.)

RATIO MATCHING

The earliest form of ratio matching appears to have been devised by Merkel (1888) in order to determine what he called the "doubled stimulus." It was a direct ratio-scaling method, but its potentialities were not effectively exploited. The method would now be classed as one of the varieties of ratio production. It is convenient to distinguish three subclasses of ratio matching, as follows:

Cross-Modal Ratio Matching

J. C. Stevens set two different brightnesses in front of the subject and asked him to adjust one of two noises to make the ratio of the noises match the apparent ratio of the brightnesses (see S. S. Stevens, 1961a, pp.

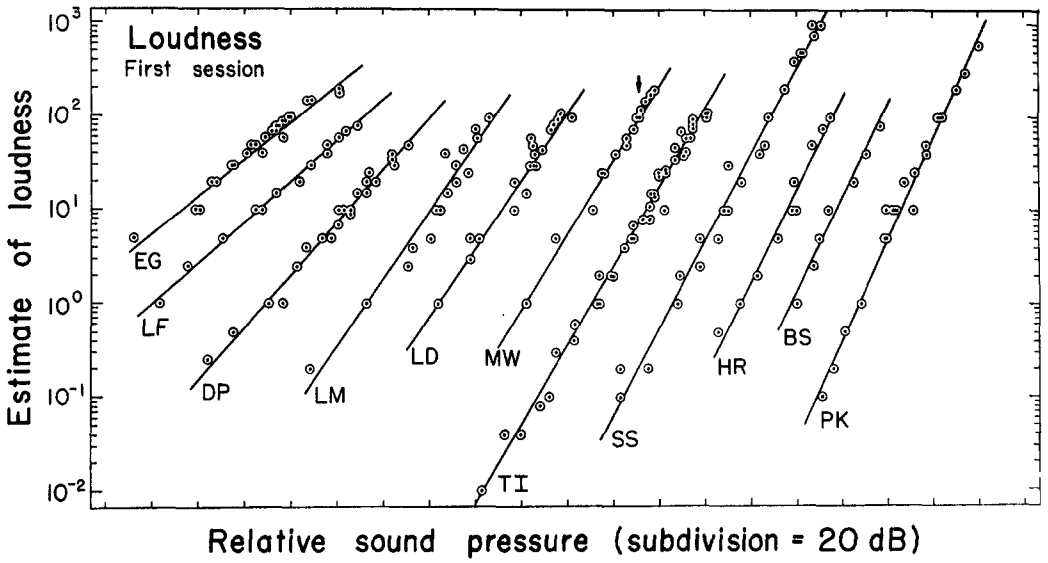

FIG. 3. Individual functions obtained when each of the 11 observers set the stimulus level and estimated the loudness. (Each point represents a judgment. There was no averaging—from J. C. Stevens & Guirao, 1964.)

17–18). The results confirmed the relative values of the exponents for loudness and brightness, showing, in fact, that the two exponents are approximately equal.

Ratio Estimation

Here the subject matches numerical ratios to apparent stimulus ratios. In the "complete" version of the procedure, stimuli are presented in all possible pairs and the apparent ratios are estimated (Ekman, 1958). Other versions use fewer stimulus pairings, and some versions provide for reporting in terms of fractions or percentages. For example, in an early experiment, Ham and Parkinson (1932) presented a sound at one level followed by a sound at a lower level, and asked the subjects to estimate what percentage of the loudness remained.

Ratio Production

In this once-popular scaling procedure, the subject is required to find or produce the stimulus that seems to stand in a prescribed relation to a standard stimulus. As we have seen, Merkel invented that kind of task with his method of doubled stimulus. Fractionation is the name commonly used for procedures that require the subject to set a stimulus to one-half (or some other fraction) of the standard. (For a tabulation of many of the numerous ratio productions that have been made with acoustic stimuli, see Stevens, 1955a.)

Ratio production has fallen into disuse mainly because magnitude matching seems to be a superior procedure. The biases in ratio production are such that the method often fails to produce a clean power function.

INTERVAL MATCHING AND VIRTUAL EXPONENTS

Although the judgment of intervals or differences, as required in various kinds of partitioning operations, may produce satisfactory results on metathetic continua, systematic biases afflict partitions carried out on prothetic continua. Furthermore, the partitioning operations can produce at best an interval scale, not a ratio scale. Nevertheless, one or another form of interval matching has produced data that have played a role in the establishment of the psychophysical power law (Stevens, 1953).

In order to describe the procedures used

and the results obtained in the various kinds of partition operations, it is convenient to distinguish two exponents: a virtual or functional exponent and the actual exponent of the continuum in question. The virtual exponent is the one the observer appears to be using when he makes his partition judgments. It is an "as if" exponent. The value of the virtual exponent α turns out to be lower than that of the actual exponent β. Since $\alpha < \beta$, the scales created by partitioning are nonlinear relative to the corresponding magnitude scales created by magnitude or ratio matching.

Perhaps the best known example of a partition scale is the Munsell scale for the lightness of grays. That scale has been determined and redetermined by several kinds of partition operations. A series of gray papers may also be scaled by magnitude estimation, as was shown by Stevens and Galanter (1957). The virtual exponent of the Munsell scale is approximately .33 and is decidedly lower than the actual exponent, approximately 1.2, obtained by magnitude estimation.

Let us now consider four varieties of interval scaling procedures.

Cross-Modal Interval Matching

This procedure seemed to work well in a 1953 experiment when subjects adjusted markers along a line (position, a metathetic continuum) in order to match the apparent spacing of a series of loudnesses. Subjects also matched marker position to the apparent spacing of the heaviness in a series of lifted weights. The same principle is involved, of course, in numerous rating scales: the subject expresses his opinion by marking a position on a line. Newhall (1950) used a somewhat similar method, involving markers on a two-dimensional grid, in order to determine spacings among the apparent lightnesses of gray papers. His results agreed with the Munsell scale. As a general method, however, interval matching suffers from a basic ambiguity, especially when metathetic position is not one of the continua used. If two prothetic continua are involved, the subject may find it easier to match ratios than differences, and without intending it, he may actually produce ratio matches between the pairs of stimuli on the two continua.

Efforts to judge intervals often encounter a dramatic hysteresis effect, which makes the judgment highly contingent on the order in which the stimuli are presented (Stevens, 1957b).

Interval Estimation

Here the subject may be asked to assign numbers to represent the sizes of apparent differences. For example, Dawson (1968) presented pairs of loudnesses and asked the observers to make a magnitude estimation of the apparent difference in each pair. He also presented pairs of visual areas. The typical biases that emerge under partitioning procedures were apparent in the results, especially in the judgments of loudness differences. A constant loudness difference (in sones) is not judged to be constant; rather a given difference is judged smaller when it is moved up the stimulus scale.

Another demonstration of the bias in interval judgments—the operation of a virtual exponent—is contained in the results of Beck and Shaw (1967) who asked 28 subjects to judge four loudness intervals, 5, 10, 15, and 20 sones in width, each located at four stimulus levels. The median estimations for three of the interval sizes are shown in Figure 4 as a function of the sound pressure level of the tone at the lower end of the interval. The curved lines show the general trend of the data.

If the world were so constructed that equal prothetic intervals appeared equal to the perceiving subject, the lines in Figure 4 would be straight and horizontal. The downward trend of the data in Figure 4 illustrates the typical result obtained in partition judgments of whatever variety: equal intervals are not judged equal at different locations on a prothetic continuum. As an interval of a constant size moves up the scale of the continuum, the constant interval is judged to be smaller and smaller.

The curves in Figure 4 were generated by a partition model in which it was postulated that the observer's judgments are governed

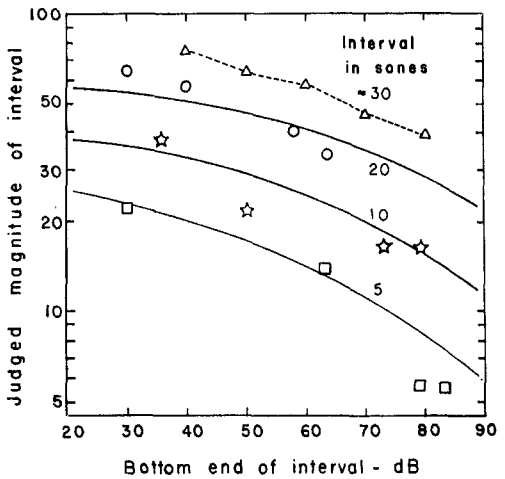

FIG. 4. Showing how the judgment of an interval of a constant size depends on the location of the interval. (Observers made magnitude estimations of sets of intervals 5, 10, 20 sones wide. In another set of experiments (triangles) the intervals were approximately 30 sones wide. The stimulus level at the bottom end of the interval is shown by the abscissa. The ordinate gives relative values only. As a constant interval moves upward in sound pressure level, the perceived size of the interval decreases. The family of three curves was generated by assuming that instead of the actual exponent of the sone scale .6, the observers used a virtual or "as if" exponent equal to .3. Triangles from Dawson, 1968; other data from Beck & Shaw, 1967.)

by a power law that does not have the actual exponent of the continuum, but rather has a virtual or functional exponent equal to .3. The fit of the curves is only fair, for the data do not provide enough information to distinguish between the family of functions generated by the virtual exponent value .3 and the family given by some other nearby exponent. If the observer's virtual exponent were .6, it would correspond to the actual exponent, and the lines in Figure 4 would then become straight and horizontal. As the virtual exponent becomes smaller, the family of curves tilts more steeply downward, and the distance between the curves decreases.

To a first approximation, then, it appears that the observer judges loudness intervals as if his power function had a virtual exponent about half as large as the actual exponent of the continuum. That principle was rather nicely confirmed in a second experiment by Beck and Shaw (1967) in which 29 observers made magnitude estimations of another set of loudness intervals—intervals that were constructed to be constant in size as determined by the lambda scale, a scale that was constructed so as to agree with a particular set of bisection data (Garner, 1954). Over the stimulus range of interest here, the lambda scale has an effective exponent of approximately .3. In other words, the "constant" intervals provided by the experimenters were generated by a function whose exponent coincided with that of the virtual exponent. As we should expect, therefore, the judged size of the intervals did not show a downward drift with increasing stimulus level. In fact, when the judgments are plotted as in Figure 4, but with *lambda interval* rather than *sone interval* as the parameter, the data describe functions that are very nearly horizontal. Thus the principle is clear: When the generating function used to set up the equal intervals has the same exponent as the virtual operating function employed by the observers in their partition judgments, then the intervals all appear equal.

It is of interest next to consider the other extreme and to ask what happens when the equal intervals are generated by a function with an exponent that is lower than the virtual exponent. Since a power function with a very low exponent resembles a logarithmic function (see Figure 5), we can examine the problem by setting up equal logarithmic or equal decibel intervals. Again the results turn out as expected. Decibel intervals appear to grow larger as their absolute level is raised. Thus a series of successive 10-db. intervals beginning at 40 db. produced the following magnitude estimations: 1.51, 1.89, 2.65, 4.14, and 9.10 (Dawson, 1968). Those values represent averages over four different experiments. They show that the 10-db. interval 80–90 db. was judged to be about six times larger than the 10-db. interval 40–50 db. If plotted in Figure 4, the judgments of 10-db. intervals would describe a curve that sweeps upward rather than downward. In other words, observers' judgments demonstrate that the virtual exponent is decidedly greater than zero, and that a logarithmic

function does not accord with interval judgments.

As a matter of fact, although Dawson did not use intervals that were exactly constant in sones, his data agree fairly well with functions having the form of those generated by a virtual exponent of about .3. The triangles in Figure 4 show the magnitude estimations of the largest sone intervals employed, ranging from 16 to 31 sones. The magnitude estimations of those intervals were multiplied by a factor in order to bring each of them to the value it would have had if all the intervals had been 30 sones. That multiplicative correction appears to do little or no violence to the data. The path depicted by the triangles in Figure 4 shows how the apparent size of a 30-sone interval grows smaller as the stimulus level used to define the bottom end of the interval increases.

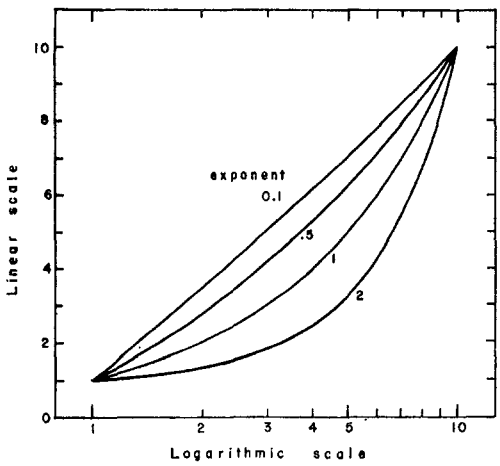

FIG. 5. Sample power functions plotted in semi-logarithmic coordinates. (As the exponent of the power function decreases, the graph of the function becomes straighter. It thereby approaches the form of a logarithmic function, which is described by a straight line in these coordinates.)

Interval Production

The production of prescribed intervals was the method invented by Plateau (1872). He asked eight artists to paint a gray such that the intervals from gray to black and gray to white would appear equal. That type of procedure is often called bisection. When more than two equal intervals are involved, it is called equisection (Garner, 1954).

The intervals to be produced need not be equal. For example, in one of my own experiments, observers produced loudness intervals corresponding to markers spaced unevenly along a line. The results from that procedure of multisection were consistent with the results obtained by equisection.

Under conditions designed to make the judgment maximally easy, the results of bisection experiments have sometimes been found to agree fairly closely with the psychophysical functions produced by magnitude and ratio matching. In other words, although the inevitable systematic difference was clearly evident, the size of the difference proved to be fairly small (see Stevens, 1955a for loudness, 1961b for brightness). To a limited extent, then, the bisection experiments have confirmed the exponents of the magnitude scale obtained by ratio-scaling procedures. The magnitude-scale exponents serve as an upper bound and the bisection-scale exponents normally lie below the bound.

In a bisection experiment in which ϕ_2 is set midway between ϕ_3 and ϕ_1 the virtual exponent of the power function can be determined by the bisection equation (Stevens, 1955a), which may be written

$$\phi_3^{\alpha} - \phi_2^{\alpha} = \phi_2^{\alpha} - \phi_1^{\alpha}.$$

The exponent α may be found by iteration. It seems to be an invariant rule that, as with other partitioning procedures, the value of the virtual exponent determined from bisection has a lower value than the exponent determined by procedures that call for ratio judgments.

The family of curves in Figure 4 tells us that a true bisection on a prothetic continuum will not appear correct to the observer. The lower half of the interval will appear larger than the upper half. Consequently, when the observer himself makes the bisection, he lowers the bisecting point. In some bisection experiments, the bisection point has fallen low enough to agree with the exponent used to generate the curves in Figure 4 (Garner, 1954). In other experiments, the virtual exponent has been found to lie closer to the actual exponent (Stevens, 1955a).

Sum Production

The virtual exponent describes the convexity in the partition function under still another procedure. The bisection problem can be turned around, so to speak, and instead of asking the observer to produce segments, the experimenter can present the segments and ask the observer to produce the whole, or the sum. If the virtual function is convex, we should expect that the perceived whole would prove to be less than the sum of the perceived parts. When observers were shown two or more line segments and asked to produce a line that appeared equal to the sum, the line produced was longer than the sum of the separate lengths (Krueger, 1970). In other words, a line equal to the sum of the separate lengths would have appeared shorter than the apparent sum. The experimental results demonstrate that the exponent 1.0, which observers ordinarily use in judging apparent length, has been replaced by a virtual exponent having a lower value.

Relation to Power Law

Perhaps the most important outcome of the work on partition scales lies in the demonstration, by means of interval methods, that the psychophysical function is a power law. The argument rests on the fact that the apparent bisection point remains approximately constant when all the stimuli are increased or decreased by the same factor. For example, Plateau's eight artists each worked in a different atelier, under a different level of illumination, and yet they all produced bisecting grays that, when viewed in the same setting, were *"presques identiques."* When that happens, the sensory magnitude must follow either a logarithmic law or a power law. If the bisection point falls at the geometric mean between the end points used to define the bisected interval, then the log function is indicated. If the bisection point falls above the geometric mean, as it seems to have done in Plateau's experiment, then a power function is indicated. And the virtual exponent α increases as a function of the distance of the bisecting point above the geometric mean. Some of those basic relations are illustrated in Figure 6.

Under optimal experimental conditions, the virtual exponent α, determined by bisection, may approach the value of the actual exponent of the continuum as an upper bound. Normally, as we have seen, the bisection exponent lies below the continuum exponent. Nevertheless, the invariance of the bisection exponent under multiplicative variations in the stimulus levels has provided evidence that equal stimulus ratios produce equal sensation ratios—which is the invariance principle that underlies the power law.

Category Scales and Thresholds

The category scale calls for special consideration because it is by far the most common and yet perhaps the least satisfactory form of partition scale. Most of the many users of category scales throughout science, education, engineering, and commerce do not intend a partitioning operation. Their hope is to grade or assess some variable, but they proceed to prescribe—and to limit!—the subjects' response scale. Commerce does much of its buying and selling with the aid of the grading of goods by subjective assessment on simple category scales, a crude procedure perhaps, but it serves its limited purpose. The business of scaling and measurement is not a primary aim in industry as it is in psychophysics. Nevertheless, category scaling is occasionally used in psychophysical experiments, despite the demonstrated inefficacy of the procedure. Stevens and Galanter (1957) examined some 70 different category scales on a dozen different perceptual continua and found little to justify the use of category scaling in quantitative studies. Their hope was that category scaling would thereafter fall into disuse. They hoped in vain.

The categories may be designated by a limited set of adjectives, such as large, medium, and small. Or the categories may be designated by a finite set of numbers, such as 1 to 6. Those were the numbers used for history's first recorded category scale, the scale of stellar magnitude, which dates from about 150 B.C. and which in a much revised form still serves the astronomer (see Stevens, 1960).

On prothetic continua, the category scale is invariably nonlinear relative to the magnitude scale. When plotted against the scale obtained by magnitude estimation, for example, the data from category judgments produce a curve that is concave downward. Whenever the subject is asked to categorize, he is forced to divide the continuum into parts or segments in order to make it conform to the limited, finite set of numbers or adjectives that he is required to use. In other words, he is obliged to attend to differences or distances. Under those circumstances, the subject is forced out of ratioing and into partitioning.

What we learn from category experiments is that the human being, despite his great versatility, has a limited capacity to effect linear partitions on prothetic continua. He may do quite well, to be sure, if the continuum happens to be metathetic, but, since most scaling problems involve prothetic continua, it seems that category and other forms of partition scaling ought generally to be avoided for the purposes of scaling. If, for some reason, an unbiased interval scale is needed, it can be obtained from a ratio scale, for the ratio scale contains the interval scale (Stevens, 1946).

The reverse is not possible, however. The ratio scale cannot be recovered from the interval scale when only interval information is available.

Is there a use for category scaling? Although essentially useless for ratio scaling, category methods play an indispensable role in threshold measurements, where the problem reduces to the determination of boundaries between classes. In fact, all threshold determinations involve category procedures, because the problem is to sort stimuli into classes, for example, those that are detectable and those that are not, those that are disturbing and those that are not, or those that are acceptable and those that are not, and so on.

Psychophysical thresholds are boundaries between classes. Although the boundary that we call a sensory threshold may be sharp at a given point in time, in a living organism the boundary behaves as though it were jittering about. Consequently, the measurement of a threshold becomes a statistical process: On the basis of repeated samples, we make a statistical decision regarding the location of the boundary. Still, the underlying experimental operation for determining any kind of threshold always involves a procedure of matching either stimulus to category or category to stimulus.

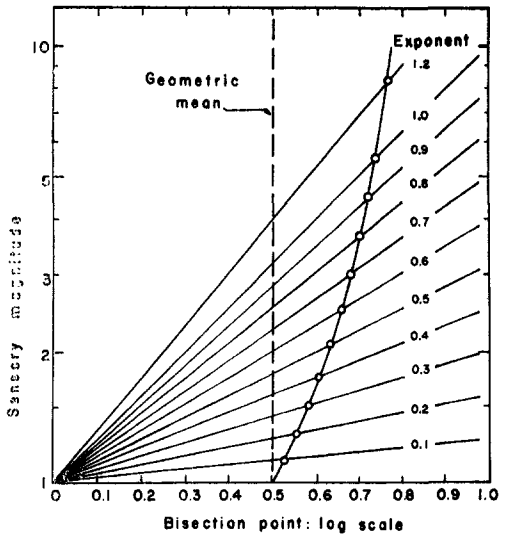

FIG. 6. Schematic diagram showing a family of power functions with exponents ranging from .1 to 1.2. (As shown by the circles, the position of the bisection point moves upward along the stimulus scale (abscissa) as the exponent increases. Bisection at the geometric mean would correspond to a logarithmic function, or a vanishingly small exponent.)

POWER-GROUP TRANSFORMATIONS

As we have seen, partition scales can often be usefully described by power functions. Consequently, the nonlinearity of the partition scale can be conveniently described by the difference between the virtual exponent of the partition scale and the actual exponent of the magnitude ratio scale. In other words, the operation of partitioning causes the observer to behave as though a power-group transformation had been performed, that is, as though an exponent had been altered. Whenever two power functions differ in exponent, they are nonlinearly related. Since the virtual exponent is always less than the actual exponent, the curvature of the parti-

tion scale, relative to the magnitude scale, is always in the same direction.

More formally, we may express the subjective magnitude ψ as a function of the stimulus ϕ by $\psi = k\phi^{\beta}$, where k depends on units, and β is the actual exponent of the continuum. The actual exponent β is the one we hope to determine more and more accurately as we learn to control regression effects and other biases. The partition scale value P can be expressed by a similar equation, but with an additive constant P_0 to take care of the arbitrary reference: $P + P_0 = k\phi^{\alpha}$, where α is the virtual exponent.

The amount by which the value of α is less than the value of β determines the curvature of the partition scale. In some kinds of equisection experiments, the curvature is so slight that the value of α has been pushed to within about 10% of the value of β. At the other extreme, in some forms of category scaling, the value of α has fallen to very low values (Marks, 1968).

It is interesting to note that the category scale for stellar magnitude cannot be expressed in terms of a power function, because the scale is more curved even than a logarithmic function. Otherwise said, the midpoint or bisection point of the visual category scale of stellar magnitude falls *below* the geometric mean of the stimulus scale. The stellar category scale is a rather special case, however, because as a stimulus array, the distribution of the stars is prodigiously skewed. An attempted representation of the stellar judgments by a power function with a negative exponent was given by Marks (1968), and a similar treatment for bisections falling below the geometric mean was given by Fagot (1963). Negative exponents, however, imply inverse or reciprocal functions and do not seem to be appropriate to the present problem.

A category production scale for loudness gave the virtual exponent .3, which is about half the value of the actual exponent for loudness. The procedure of category production serves to diminish the effects of stimulus spacing and thereby to approximate the pure form of the category scale. It is interesting that the pure category scale should have approximately the same virtual exponent as that produced by the magnitude estimation of differences, as in Figure 4.

In category scaling, the difference between the virtual exponent α and the actual exponent β seems also to depend on variability, or on the noise load imposed by the task. Some continua are easier to judge than others. For example, the curvature of the category scale was shown to increase and the virtual exponent to decrease as the continuum was changed from length of lines to largeness of squares to loudness of tones (Stevens & Guirao, 1963). The variabilities (standard deviations) with which the observers set values on those three continua under the procedure of magnitude production were: length, 1.0; largeness, 2.1; and loudness, 4.0 decilogs.

It is especially important to note that although partitioning produces a power-group transformation that lowers the effective value of the exponent, the resulting virtual exponent α is always positive. We must, of course, exclude from that generalization the category scale obtained with highly skewed distributions of stimuli, because those abnormalities are essentially artificial and can be remedied by straightforward procedures of experimental iteration (see Pollack, 1965). The iterated pure category scale seems always to have a positive exponent. The importance of a virtual exponent that is positive and decidedly different from zero lies in the evidence it provides that the category scale is not a logarithmic function of the magnitude scale. A logarithmic category scale has often been assumed (e.g., Torgerson, 1961), but under that assumption, the virtual exponent would lie near zero.

The term power group was the name proposed for the group of permissible transformations on what I called a logarithmic interval scale (Stevens, 1957b). The power group provides that any scale value x may be replaced by x' where $x' = ax^{b}$. The transformation preserves the equality of ratios, but not of differences. For the visually minded, it may be helpful to note that a power-group transformation changes the curvature, or, in log-log coordinates, the slope of a function.

It is remarkable how widespread are the

instances of power-group transformations. Figure 2 shows a dramatic example: The change from estimation to production produced a new power function with a larger exponent, hence a different slope in log-log coordinates, which means a different curvature in linear coordinates. How is such a complex power transformation produced with such apparent precision? The experimenter reverses the procedure, whereupon the observers alter the curvature of their response function in just such a way as to preserve ratios. How is that possible?

Many other circumstances produce power transformations. Among them we find the effect of adaptation on visual brightness. The exponent rises from .33 to .44 when the state of adaptation is changed from dark to about 100 db. above threshold (J. C. Stevens & Stevens, 1963). Even more drastic changes in the exponent—alterations by a factor of three or more—may take place under inhibition, which includes visual contrast and auditory masking (Stevens, 1966). Thus it appears that changes in the behavior of sense organs under changing states of adaptation, and under the inhibition created by glare and masking, can be described by power-group transformations. Similar transformations occur in the dramatic summation of warmth when the irradiated area of the skin is increased. The exponent decreases by a factor of about 2 when the area is increased by a factor of 10 (J. C. Stevens & Marks, 1971). Although I once regarded the power group as something of a curiosity and useful only in the discussion of scaling theory, the piling up of examples has led to the suggestion that the power group may prove to be one of the most common transformations in the biological domain.

A review of research since the 1930s (Stevens, 1970, 1971) turns up many experiments in which electrical recordings of neurelectric effects in sensory receptors, nerve fibers, and neural complexes have exhibited a power-law dependence on stimulus intensity. Thus, there exists rather direct evidence that sensory systems are capable of power-group transformations. Of course, every placement of an electrode does not produce a nice clean power function, but power functions have been recorded by one or more investigators in several sense modalities: vision, hearing, taste, touch, kinesthesis, and electrical pulses to the skin. The point to be made here is that there appears to be a tendency for the neurelectric exponent to decrease as the recording site becomes more remote from the sense organ. That is to say, following the power-law transduction performed by the sense organ, there may ensue subsequent power-group transformations higher in the nervous system. But in the present state of knowledge, such principles remain as vague as the evidence.

THE PARTITION PARADOX

The tendency of observers to make partition judgments as though their effective or virtual exponent is lower than the actual exponent of the continuum has impressed many people as paradoxical. Krantz (1970) wrote:

> One of the long-standing puzzles is why the scales obtained from category rating differ from those of magnitude estimation [p. 40].

Why, in other words, does a constant interval on a prothetic continuum such as loudness not appear to remain constant when the stimulus level is raised?

Perhaps no answer will satisfy everyone, but we can at least rule out the nature of the observer's task. We cannot blame partitioning as such, because partition judgments lead to linear results on metathetic continua. For example, a 5-mel interval sounds the same size regardless of its location on the pitch continuum. But a 5-sone interval does not sound the same size at different locations on the loudness continuum (see Figure 4). Hence the problem must have to do with the nature of the continuum. Among the features that distinguish the two kinds of continua, there is one aspect of prime importance. On a metathetic continuum, such as pitch, the error distribution, as measured by the size of the just noticeable difference (jnd) in subjective units (e.g., mels), remains the same all along the continuum. In other words, the *absolute* error is constant. On a prothetic continuum it is the *relative* error that stays constant, so that the jnd for

loudness, measured in sones, grows very large as loudness increases.

In order to illustrate how unlikely, indeed impossible, it would be for an observer to make a correct judgment of a loudness difference regardless of stimulus level, let us consider the following paradigm. We will assume that the loudness exponent is .6 and that the interval from 40 to 50 db. corresponds to the subjective interval between a loudness of 1 sone and a loudness of 2 sones, or a difference of 1 sone. That happens to be a clear and obvious difference. Now consider the same 1-sone difference at 100 db. where the loudness is 64 sones. An increase from 100 to 100.2 db. makes a 1-sone difference, raising the loudness from 64 to 65 sones. But a change as small as .2 db. is so small that it would be detected less than half the time. It seems beyond possibility, therefore, that a constant difference of 1 sone could be judged to be the same size regardless of where it occurred on the continuum.

That illustrative example may help to suggest why it is that every variety of partitioning must give a distorted result on a prothetic continuum. A constant difference transforms itself from obvious to undetectable as we go from weak to strong stimuli. Therein lies the essence of the prothetic principle.

A graphic illustration of the underlying principle is shown in Figure 7. The ordinate shows the stimulus change in decibels that corresponds to a change of 1 sone. As in the example above, at a level of 40 db. a change of 1 sone corresponds to 10 db. At 100 db., the required change is only .2 db., and it becomes still smaller at higher levels. The jnd (Weber fraction) is plotted in Figure 7 as a horizontal line with an ordinate value of about .5 db. (see Miller, 1947). A plot similar to Figure 7 can be constructed for any prothetic continuum. For visually presented lengths, for example, such a plot could be made to show how, in visual perception, a centimeter added to a centimeter makes a clearly perceived difference, whereas a centimeter added to a meter becomes only marginally detectable. Under successive visual presentation, length behaves like loudness, and a constant added difference becomes lost to view as the stimulus increases.

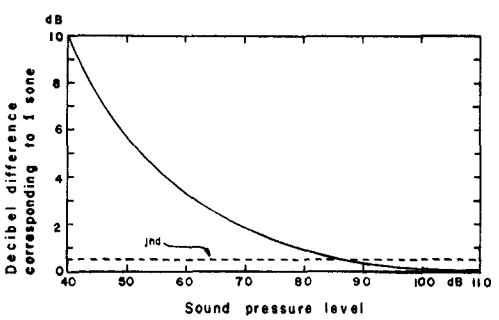

FIG. 7. Showing how the difference in decibels required to produce a 1-sone loudness difference falls off with increasing sound pressure level. (At 40 db. it requires an added 10 db. to add 1 sone. At 110 db. it requires an added .1 db. to add 1 sone. The dashed line shows the approximate value of the jnd, the increase that would be detected about half the time.)

The smooth continuity of the curve in Figure 7 suggests that the underlying process that forces a given constant difference to become less and less apparent as the level increases is a process that operates all up and down the prothetic continuum. There are no discontinuities, no sudden transitions. When the observer attempts a comparison of differences at any place on the continuum, it is as though his perception undergoes an asymmetrical distortion; a constant difference seems larger toward the lower than toward the higher part of the continuum. From that basic asymmetry, it follows that the operations of partitioning on prothetic continua will fail to produce a linear, unbiased interval scale. Where the asymmetry does not exist, as on a metathetic continuum, it becomes possible for partitioning to produce a linear interval scale.

INDIVIDUAL DIFFERENCES

Let us turn at this point to a problem that was especially well formulated by Jones and Marcus (1961). Do different individuals have different operating characteristics in their sensory systems? If not, then why do the slopes (exponents) differ when observers match numbers to stimuli, as, for example, in Figure 3?

If we were to take some of the measured values literally, it would be incumbent on us

to examine the implications. Consider in Figure 3 the functions for observers EG (exponent .4) and PK (exponent 1.1). They both happen to be "normal" observers, having similar audiometric thresholds. If, as seems reasonable, the two observers experience the same loudness when a tone is near threshold, what about a tone 100 db. above threshold? The two exponents suggest that the loudness at 100 db. would be about 3,000 times greater for PK than EG. Yet both observers react very similarly to acoustic stimuli. For example, both observers report that 100 db. sounds very loud; they both call 120 db. slightly painful; and they both jump when it is first turned on. Both observers set the intensity control of the radio to roughly the same level for comfortable listening. Both carry on conversations in about the same level of voice. All those bits of evidence make it hard to credit a difference in exponent that entails a 3,000-fold difference in loudness at 100 db.

Another consideration is this. The value of an exponent reflects the curvature in the operating characteristic of the sensory system. As a function of sound pressure, the exponent .4 suggests that the loudness function for EG is a decelerating function, sharply concave downward. The exponent 1.1 suggests that the function for PK is concave upward—an accelerating function of sound pressure. In order to achieve such an unlikely difference in curvature, nature would have had to produce two radically different kinds of operating characteristics in the two auditory systems.

An alternative hypothesis is that all human auditory systems operate on much the same design, with very nearly the same exponents, but that there are wide individual differences in what observers take to be the numerical value of a loudness ratio. We encounter that kind of individual difference especially clearly when an observer is asked to adjust one sound to make it appear half as loud, or twice as loud, as another sound. In an experiment with 22 observers, the scatter of eight distributions of settings covered a median range of 14 db. (For plots of the distributions, see Stevens, 1957a.) That range of settings would correspond to a range of exponents from about .3 to 1.5. The distributions of the ratio settings tended to be roughly log normal, however, and there appeared to be no reason not to average the data, which is equivalent to averaging individual exponents.

Consider this question: What would happen if an experimenter were to report that each of 22 observers in a group made an identical setting when asked to produce a two-to-one loudness ratio? It is an interesting question whether the scientific community would regard such a result as a new marvel, or whether scientists would simply reject it as an implausible outcome. We expect variability and we usually find it. The substantive question concerns whether and how to average the data.

Although the operating characteristics of normal sensory systems may have the same or closely similar exponents, there are circumstances in which an abnormal exponent can be demonstrated and in which the abnormality calls for careful measurement. The otologist, contemplating middle-ear surgery, needs to know whether so-called recruitment, with its attendant large exponent, characterizes the lower part of the patient's loudness function. If the exponent is abnormally large, middle-ear surgery is contraindicated. But if the exponent is normal, the otologist may feel free to try to correct a middle-ear difficulty. (For a model depicting the manner in which Ménière's disease may produce power-group transformations on the loudness function, see Stevens & Guirao, 1967.)

Except for clinical cases, it seems fair to say that seldom, if ever, have investigators undertaken the serious effort that is needed to establish the existence of individual differences in sensory power functions. Mostly, we are shown the variability that happens to be found in a particular matching response, usually magnitude estimation. The obvious next step, usually not undertaken, would be to try magnitude production. I have found that many individual differences vanish as soon as the matching task is inverted. For example, by magnitude estimation, the lowest loudness exponent in a group was .4, but by magnitude production, that particular ob-

server produced the exponent .9; the geometric mean of the two exponents is .6 (see graphs in Stevens & Greenbaum, 1966).

Five observers in those same experiments (Stevens & Guirao, 1962) made both magnitude estimations and magnitude productions of the loudness of noise. The ratio of the largest to the smallest exponent for estimation was 1.42; for production, it was 1.27. When the geometric means of the exponents for estimation and production were examined, the range ratio fell to 1.14. In other words, the individual differences were less pronounced when both estimation and production were used in a balanced design.

The balancing of estimation by production may be good for a start, but if we are seriously interested in the power function for a particular individual, we will not stop with magnitude estimation and production. We will want to know the results from a balanced array of additional cross-modality matching tasks, the more the better.

VARIABILITY AND AVERAGING

Criticism of the power law has sometimes centered on one or another aspect of variability. It would be good, of course, if variability could be reduced, so that the psychophysical functions could be determined with higher precision. But empirical functions always suffer from variability, and the central question is not so much whether a measurement is variable, or whether subjects disagree, but whether averaging is appropriate. If the data can be appropriately averaged, it does not matter how widely the variability may range, provided the number of independent measurements can be increased. In principle, the standard error can then be brought down to any desired level.

In electrophysiology, for example, miracles of averaging are performed routinely by computers programmed to dig a particular waveform out of the myriad variations in the ongoing neural activity of the brain. A repeated click delivered to the ear can then be seen as an evoked potential at the scalp. The response to the click emerges remarkably clear and unencumbered by the noise of the brain, for the noise has been suppressed by being averaged out. An analogous strategy of error cancellation by averaging finds usefulness in psychophysics and in all the rest of science.

In the long run, since scientists tend to believe only those results that they can reproduce, there appears to be no better option than to await the outcome of replications. It is probably fair to say that statistical tests of significance, as they are so often miscalled, have never convinced a scientist of anything. By contrast, a tabulation of 178 determinations of the loudness exponent, based on 25 years of accumulated results from several different laboratories, produced a median result .6, which became the exponent recommended by the International Standards Organization (see Stevens, 1955a). Since then, a further accumulation of experimental determinations has begun to fix the second decimal place, and it now appears that the value 2/3 may be more representative (Stevens, in press). The value 2/3 happened to correspond to the modal value of the 1955 distribution, but at that time, the median seemed a better choice than the mode.

How to average data presents serious and interesting questions. The median is perhaps the single most unbiased measure of location, and it has often been used in psychophysics. A popular rule is: when in doubt, use the median. On the other hand, a more efficient average is often wanted, and the choice of an efficient measure can usually be made to rest on the form of the distribution. Thus, two different averages have proved appropriate in psychophysical scaling, each under a particular circumstance.

In an equisection experiment, 45 subjects divided a 40-db. segment of the loudness continuum into four equal-appearing intervals. Because the decibel measures of the subjects' settings gave skewed distributions, it did not seem proper to average the decibel values, which would have been equivalent to computing geometric means. When the decibel measures were converted into sones (a linear loudness value), the settings showed the desired symmetries. It was concluded that averaging should be done by computing the arithmetic means of the loudness values, and an iteration procedure for determining those

values was outlined (Stevens, 1955b). The arithmetic mean is also an appropriate average for other kinds of partition scales, such as the category scale.

In experiments involving magnitude estimation and other forms of cross-modality matching, it is the geometric mean, not the arithmetic mean, that appears to be the appropriate average. Against a linear scale, the response distributions are skewed, as indeed they must be when error is proportional to magnitude. When error grows in proportion to magnitude, so that the relative error stays constant, a logarithmic transformation tends to undo the skewness. The applicable principle states that when error is relative, the error distribution is log normal.

The log-normal model for magnitude estimation was tested by J. C. Stevens who plotted the results obtained when 70 naive observers estimated the loudness of a white noise presented by a loudspeaker in a classroom (J. C. Stevens & Tulving, 1957). In the first of two experiments, the observer chose his own modulus by assigning to the first stimulus (85 db.) whatever number seemed appropriate. In the second experiment, the stimuli were presented in pairs, a standard at 85 db. called 10 followed by a variable. When the cumulative frequencies for this second experiment were plotted on probability paper against the logarithm of the magnitude estimations, the result was a family of straight lines. In other words, the distributions were log normal.

Some five years later J. C. Stevens and Miguelina Guirao applied the modulus equalization procedure to the data of the first experiment, the one in which each of the 70 subjects had chosen his own modulus. (The deviation of each subject's scores from the average function was minimized by the procedure for modulus equalization described above.) The cumulative frequencies of the judgments subjected to modulus equalization are shown in Figure 8. Again the straight lines demonstrate that the distributions are approximately log normal.

Other features of the results in Figure 8 are also of interest, especially in view of the fact that those were the first magnitude estimations ever made by that large group of

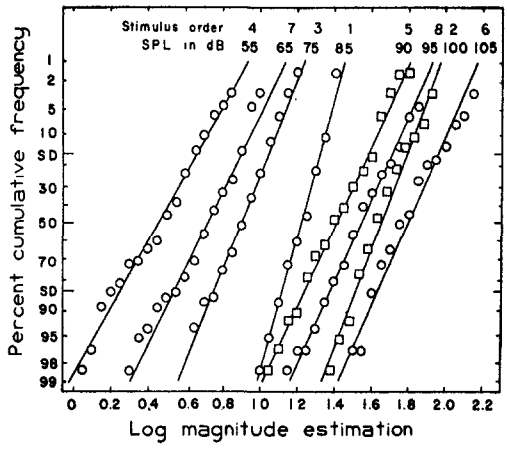

FIG. 8. Cumulative frequency distributions of magnitude estimations. (Each of 70 subjects, making their first judgments ever, assigned whatever number seemed appropriate to eight levels of white noise presented in the order shown. The data of three subjects who used negative numbers or zeros were not tabulated. In these coordinates, a straight line signifies a log-normal distribution. Data from J. C. Stevens & Tulving, 1957.)

naive listeners. Since it was a classroom experiment, the order of the stimuli could not be made different for the different listeners. Consequently, the order of the stimuli is reflected in the slopes of the cumulative frequency lines, or, in other words, in the standard deviations. The first stimulus has the smallest standard deviation (1.0 decilog), the second stimulus the next smallest (1.4 decilog), and so forth. If the stimulus order had been made different for each observer, the lines in Figure 8 would be more nearly parallel. But a tendency has been observed in numerous experiments for the variability to increase at the low end of the scale, and to a lesser extent, at the high end. That same tendency is apparent in Figure 8. Hence, it can be seen that the effect due to stimulus order cuts across the normal tendency for the variability to be slightly lower in the middle range.

The lines in Figure 8 make it clear that the geometric mean is an appropriate average for the data. In this instance, the geometric means determine a power function with the exponent .55.

A similar treatment was applied to the data from the original "no standard" experi-

ment carried out in 1954. The elimination of the standard was an important procedural change suggested by Geraldine Stevens. A group of 32 observers each made two judgments of each level of a 1000-hertz (Hz.) tone. (The details of the procedure are given in Stevens, 1956.) The level of the first tone was varied from one observer to another and was assigned whatever number the observer thought appropriate. The cumulative frequencies were nicely log normal, both before and after modulus equalization. The effect of the modulus equalization was to reduce the standard deviations by a large factor, as shown in Table 1.

It is significant to note that the standard deviations were larger when the stimulus to be judged was 1000 Hz. (Table 1) than when it was a white noise (Figure 8). In numerous experiments, it has been found that a noise is easier to judge than a tone.

Still more difficult for most observers are magnitude estimations of apparent pitch. Each of 20 subjects made two judgments of 12 frequencies between 100 and 7500 Hz., with no designated standard (Stevens & Galanter, 1957). With such a small number of scores, the cumulative frequency plots showed much scatter, but the overall picture was log normal. Corrected by modulus equalization, pitch judgments yielded standard deviations that were not very different from the corrected values for the loudness judgments in Table 1. For pitch, the standard deviations ranged from 1.7 to 5.4 decilogs, with a median value of 2.2 decilogs.

TABLE 1

STANDARD DEVIATIONS IN DECILOGS OF MAGNITUDE ESTIMATIONS DETERMINED GRAPHICALLY FROM CUMULATIVE DISTRIBUTIONS

Stimulus	*SD* uncorrected	*SD* corrected
110	5.6	3.0
100	5.4	2.3
90	4.7	1.8
80	4.7	1.7
70	4.6	1.7
60	5.0	1.9
50	5.5	3.0
40	5.8	3.2

Note.—Thirty-two subjects made two judgments of each stimulus (1000 Hz.). When corrected by modulus equalization, the variability fell by a factor of approximately 2.

The variability in the results of cross-modality matching for a group of observers can be divided into three main components, namely, the variability attributable to differences from observer to observer in the effective modulus (intercept), the effective exponent (slope), and the residual scatter in each observer's matches. It is sometimes useful to partial out those sources of variability (see Stevens & Stevens, 1960).

RANGE EFFECTS

Much has been written about the effects of stimulus range on the exponents of the sensory power functions. The range (logarithmic spread) of the stimuli used in cross-modality matches may affect the exponent, but the experimenter can design tactics to offset the biasing effects of range, if he so chooses. Analogous options face the experimenter with regard to other distorting factors, and the same principle applies to measurements in physics as well as psychophysics. A scientific measurement of serious consequence can never be based on one experiment, because multiple experiments are required to detect and correct the systematic errors. Multiple experiments are the rule in physics; they ought also to be the rule in psychophysics.

There is a negative correlation between the measured exponents and the ranges of the stimuli that were used in some of the experiments by which the exponents were determined. For example, a range as great as 90 db. has been used for loudness, which has a low exponent, compared with a range of about 10 db. for electric current through the fingers, which has a high exponent. The negative correlation between range and exponent has led Poulton (1968) to say

> that in designing the experiments to measure the exponent, the experimenters did not adequately compensate for the effects of the different physical ranges . . . [p. 5].

To be sure, the experimenter could choose a 10-db. range for the study of loudness; but a 90-db. range of electric current through the fingers would prove insupportable. It seems

that stimulus ranges are to a very large extent selected by experimenters because nature's exponents are what they are, not the other way around.

In suggesting that the experimenter should compensate for the effects of the different physical ranges, Poulton directed attention to the wrong side of the equation. It is not the physical ranges that need compensation; rather, the experimenter should try to ensure that the subjective ranges are as comparable as possible. Stimulus measures have much arbitrariness about them: measures of sound pressure give one loudness exponent; measures of sound power give an exponent that is half as large. For apparent size, the measured diameter of circles gives one exponent, the measured area of circles gives another, and so on.

In comparing the exponents of different continua, the experimenter would like to be able to select stimuli—regardless of how they happen to be measured—so that they would produce a constant subjective range. If he could do that, then the correlation would uniquely fix the relative values of all the exponents. Of course, if the experimenter knew in advance how to choose the stimulus ranges that would produce the perfect correlation, he would not need to run the experiment. In effect, then, much of the extensive work with cross-modality comparisons can be regarded as an effort by trial and error to determine what stimulus ranges would be needed to provide a constant subjective range on all continua and thereby make it possible to produce a perfect negative correlation between logarithmic range and exponent.

The correlation between stimulus range and exponent has been reported as high as −.94 by Teghtsoonian (1971), who proposed

> that a single scale of sensory magnitude serves a wide variety of perceptual continua, and that variation in power law exponents is primarily due to variation in dynamic ranges [of stimuli] [p. 71].

That is an interesting hypothesis, even though a test of it would require that we learn how to determine dynamic range, which may prove to be an elusive variable.

It is important to note that a short range does not necessarily produce a large exponent. Some of the lowest measured exponents apply to odor. Benzaldehyde (synthetic almond) gave the exponent .2 (Stevens, 1957b). That was probably the first olfactory exponent ever determined. A similar low value has since been found by Berglund, Berglund, Engen, and Ekman (1971). The stimulus range for benzaldehyde is relatively short, at least as compared to loudness or brightness. From the point of view of the observer, the subjective range of the odor seems extremely short compared to the enormous subjective ranges that can be produced in loudness or brightness.

EFFECTS OF REPETITION

Although range effects may be present in any given experiment, the degree to which they affect the outcome may sometimes be altered by repeated presentation of the stimuli.

An experiment showing how repeated presentations of a very short range of luminous targets can cause the measured brightness exponent to increase on successive presentations was carried out in 1960 by A. W. F. Huggins (reported in Stevens & Stevens, 1960). The results are shown in Figure 9. The stimuli, a series of Munsell grays ranging from black to white, were viewed under so-called reduction conditions, which made them appear as luminous targets, not as surfaces. The stimulus range covered only 16 db., because that is all there is between a black paper and a white one. The stimuli were also presented under two levels of illumination, which extended the range and which gave the filled points in Figure 9. The exponent for the extended range is .35. The magnitude estimations for the shorter range on the first presentation follow very closely the lowest five points on the extended range, but later presentations give successively steeper slopes (higher exponents). Although the limiting of the experimental procedure to a single presentation of each stimulus may attenuate some of the effects of range, residual effects on the exponent may still remain. Under some circumstances, the residual effects can be balanced out in the experimental design, as is shown below.

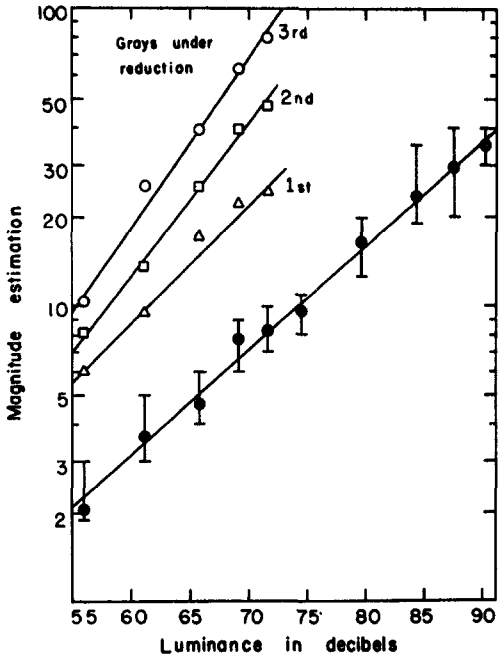

FIG. 9. Showing how Munsell grays, extending from black to white, are judged when they are presented entirely alone with no other visible luminance in the field of view. (The luminance then covers a range of about 16 db., which seems subjectively rather short. On the three successive presentations of the set of stimuli, the exponent became larger (unfilled symbols). The filled symbols show the results for an extended luminance range. The exponent is .35, which is much lower than the value 1.2 obtained when the Munsell grays are placed one by one on a table in front of the observer.)

COUNTERBALANCING FOR RANGE

The general problem of the interaction of range and exponent may be thought of in terms of the matching of two continua, A and B, each of which may serve in turn as the adjusted and as the criterion continuum. What does the observer tend to do to the adjusted stimulus as a function of the range (logarithmic difference) between two criterion stimuli set by the experimenter? For purposes of discussion let us assume that the true exponent determined by the slope A/B in log-log coordinates is 1.0.

If the experimenter were to set the criterion range to a very small value, say, to a small fraction of a decilog, on the average the observers would necessarily respond with adjusted ranges that average larger than the criterion range. That value of the adjusted range would then determine an exponent greater than 1.0. On the other hand, if the criterion range was set at a very large value, the observers would tend to match it with shorter ranges, determining thereby an exponent smaller than 1.0. Between those two extremes, the exponent would decrease monotonically as the range increased.

Since the two continua A and B can be interchanged in their roles of criterion and adjusted variable, in a balanced experiment two paths would be traced by the exponent as the criterion range was varied from very small to very large. Those two paths would cross, and the crossing point would presumably determine the unbiased exponent—unbiased by the effect of range, but not necessarily free of other possible biases.

A concrete example of the two paths followed by the exponent when range is varied in a fairly well-balanced experiment can be constructed from the data of Stevens and Poulton (1956). The loudness function for a 1000-Hz. tone was studied by allowing groups of 8 to 11 unpracticed observers to make only a single judgment, either a ratio estimation or a ratio production. A standard stimulus was sounded first and called 100, and the observer expressed the apparent ratio by assigning a number to a second stimulus at a lower level. The exponents corresponding to the median estimations of each group are shown by the triangles in Figure 10. Each triangle represents a different group of listeners.

For the ratio productions, the observers adjusted a sone potentiometer to produce a prescribed fractional loudness relative to a standard. The exponent corresponding to the ratio productions (sone average) for each group of subjects is shown by a circle in Figure 10. Again each circle represents a different group of listeners. The data show an approximate symmetry and suggest that for those naive observers making their first loudness judgments, the exponent lies between .6 and .7. The range effect emerges as a dramatic but orderly variable, and since it shows an approximate symmetry, it can in

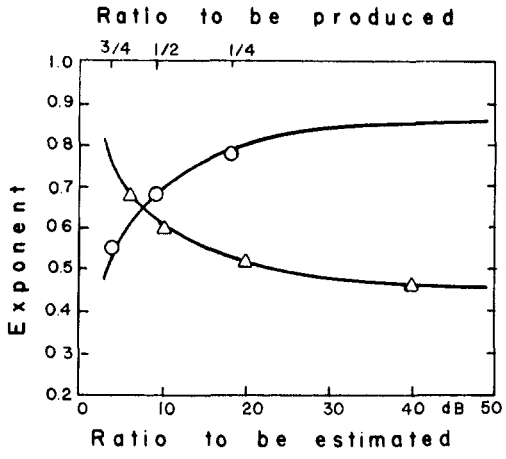

FIG. 10. The range effect in a partially balanced experiment. (Each point represents a separate group of subjects who estimated a loudness ratio (triangles) or produced a loudness ratio (circles). When the range effect produces symmetrical functions, as is approximately true here, the exponent may be uniquely determined. It is the value that makes the exponent corresponding to the crossover point equal to the exponent implied by the relation of the scales at the top and bottom of the figure. The exponent so determined is .65. Data from Stevens & Poulton, 1956.)

principle lead to a unique exponent determined by the crossing point in Figure 10.

The actual crossing point depends, of course, on the relation between the two scales (top and bottom in Figure 10) against which the two sets of data are plotted. But the relation between the values on the two scales (both logarithmic) also expresses an exponent. The problem, then, is to adjust the relation between the two scales (ratios to be produced and ratios to be estimated) so that the exponent determined by the scale relation coincides with the exponent determined by the crossing point. The adjustment can be carried out by iteration.

In plotting Figure 10, I first assumed that the exponent was .60. Accordingly, the ratio 1/2 at the top of the graph was set directly above 10 db. at the bottom of the graph. The other ratios were then set according to a logarithmic spacing. When the data were plotted, the crossover was found to correspond to .64 on the ordinate—a value that was larger than my assumed exponent. I next assumed a larger exponent, .67, and changed the upper scale and the location of the circles accordingly. The resulting crossover then fell at .65—a value smaller than my assumed exponent. Thus the two assumptions, one too high and one too low, had succeeded in bracketing the exponent between .64 and .65, but closer to .65. With fallible empirical data, the exact value of the exponent cannot be taken too literally, but it is nevertheless interesting that the crossover value accords approximately with the consensus of other measurements.

The search for other instances in which a pure range effect could be studied has been only partially successful. Experiments with the required balanced design seem to be rare, but we can compare two separate experiments that add up to a partly balanced design. In Figure 11, the triangles show the exponents corresponding to the median ratio estimations made by groups of about 30 subjects (Poulton, 1969). Each group judged one noise level set at 5, 20, or 35 db. above a standard noise (600 to 1200 Hz.) at 65 db. The circles in Figure 11 represent the exponents determined from the ratios produced

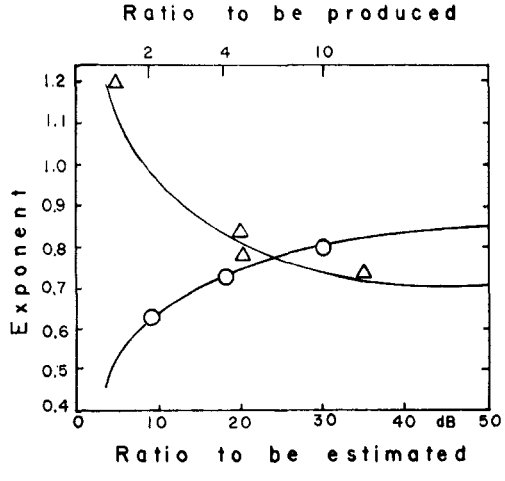

FIG. 11. Range effect shown in two noncomparable experiments. (Separate groups of subjects estimated loudness ratios (triangles) of a band of noise, 600–1200 Hz. (Poulton, 1969). A single group of subjects produced fractional and multiple ratios (circles) of a 40-tone complex (Geiger & Firestone, 1933).)

by 31 listeners who made multiple and fractional settings of noise consisting of a 40-tone complex (Geiger & Firestone, 1933). Both the multiple and the fractional productions were averaged to determine the exponents. The same group of listeners heard all the stimuli, and the order of the ratio productions was 1/2, 1/4, 1/10, . . . 2, 4, and 10.

In order to plot the circles in Figure 11, the exponent was assumed to be .67, and the ratio 2 on the upper scale was set directly above 9 db. on the lower scale. The crossover point then fell at .77. If the exponent is assumed to be .8, and if the scale values and the circles are moved accordingly, we find that the crossover point becomes about .78. Thus the two assumed exponents have bracketed the crossover exponents. Consequently, the exponents that would be determined by the data in Figure 11, if they were homogeneous and suitable for such a purpose, would lie between .77 and .78. That value is on the high side, even though the measured exponents for noise are often larger than those for tones. The mean exponent for white noise in an earlier compilation was approximately .72 (Stevens, 1955a).

The groups of subjects who gave the results that determined the triangles in Figure 11 subsequently judged all the other stimuli. The pooled results determine a power function with an exponent of about .73, provided no attempt is made to force the best fitting line to pass through the standard stimulus. Although experimenters have sometimes assumed that the function must pass through the standard, there is in fact no such requirement. The observer does not respond to any stimulus, whether standard or variable, with zero error.

The foregoing examples of the sources of error and distortion that may plague a psychophysical measurement are more illustrative than exhaustive. As has often been said, there are many ways to perform a bad experiment. Not even the expert in statistical design can tell exactly how to perform a good one. In principle, however, we can study each experimental ailment and utilize the rules of its behavior in order better to diagnose the nomothetic substrate.

THE NOMOTHETIC IMPERATIVE

The scientist's contest with nature has prospered to the degree that simplicities and uniformities have been detected amid the complexities that afflict observation and experiment. The simple invariances have often proved hard to find, however, because no experiment can be performed without its "context," and no measurement can be made without error. In psychophysics, each experimenter records results that disagree to some extent with those of his colleagues, and a penumbra of uncertainty surrounds even our best determinations. Consequently, there is room for many hypotheses and for many views regarding the structure of the psychophysical domain.

First there are those whose working hypothesis states that there exist laws to be discovered. Heeding the nomothetic imperative, those investigators refuse to be put off by the apparent chaos in the organism's reactions to stimuli, and they try to order, classify, and systematize the behavioral facts. Some people seem willing to gamble that the sense organs operate in beautifully simple ways and that what we take to be complexity lies more in our inept descriptions than in nature's actual comportment. As regards the particular question of reactions to stimulus intensity, the nomothetic outlook assumes that there exists an orderly input–output function, a simple law of some sort—however difficult it might be to pin it down in its exact form. Perhaps it was that same nomothetic outlook that drove Kepler to search for simple invariant laws in the baffling paths of the wandering planets, which were thought in early times to crisscross the skies impelled by their own volition.

The nomothetic viewpoint is consonant with an objective, operational approach, but it does not necessarily entail a particular philosophy of reality. In the fitting of schematics to empirical fact—what I have called the schemapiric endeavor—there is no necessity that we take a stand regarding any ultimate concern. We can simply note, for example, that when the intensity of the visual stimulus increases, the observer's response changes in an orderly fashion. If he is in-

structed to squeeze a hand dynamometer to match the apparent intensity of what he sees, then the force of his squeeze increases as a power function of the physical luminance (J. C. Stevens, Mack, & Stevens, 1960). That is the first-order fact. But what shall we say about it? We have obviously measured something, but have we measured sensation? As soon as that word appears, a discordant chatter sets in, because many authors contend that sensation is not a thing that can be measured.

If I thought it would help, I should happily give up the word sensation in favor of some other term, such as behavioral response or apparent effect, but the philosopher would shortly discover that my new term is only a euphemism for what I regard as a straightforward construct, a construct that can perhaps best be communicated to other people by my labeling it sensation.

A prime source of confusion in this semantic issue rests with the fact that each of us experiences sensations, and to each of us our sensations seem personal and private and inaccessible to measurement. As scientists, we should try to ignore that fact. To help free us of narcissistic introspection, there is the example of Plateau, the blind physicist, who gave us our first interval scale for the lightness of grays. He himself did not need to see the grays that the eight artists painted, for he could record the reports of other observers. Plateau's achievement drives home a crucial point: the blind could develop the psychophysics of vision; the deaf could develop psychoacoustics.

Sensation then is a construct, a name given to a constellation of behaviors. The justification for saying that sensation is subjective is that a human subject exhibits the behaviors. The heating coil in an electric stove also exhibits interesting behaviors, but quite arbitrarily we refrain from using the term subjective for those behaviors or for the constructs built on them. The simple, operational dichotomy into subjective (meaning people) and objective (meaning not people) seems eminently convenient, but words and slogans sometimes cause clash and conflict. In the schemapiric view of science, words and symbols serve only the neutral purpose of implementing a schematic structure which may be related by operational rules to an empirical structure (see Stevens, 1968b).

As an experimenter, I feel free to use terms like sensation and subjective, because they can be defined operationally in the context of psychophysical research. For some writers, however, the use of such words puts the user in the camp of Subjectivism, as opposed to Behaviorism (e.g., see Baird, 1970). I have long thought of myself as a behaviorist, but it has not seemed defensible to assert that verbal taboos provide a tool for advancing science, as some behaviorists have seemed to believe. A viable injunction is this: Use any words you care to, but let it be plain what operations lie behind your verbal forms. Otherwise said, construct the schema as you will, but tell us precisely how the empirics articulate with the schematic terms in order to produce the schemapiric substance.

Such protestations, I realize, will avail nothing with those who want to maintain a philosophical distinction between certain traditional views and to preserve a dualism that to an operationist is devoid of meaning. Since I do not believe in the usefulness of the distinction, I am happy to read that

> It is impossible to locate Stevens' view precisely, since he constantly shifts back and forth between a behaviorist and some other way of treating the question . . . [Savage, 1970, p. 390].

More damning than that, however, the whole psychophysical enterprise is said by Savage to be wrongly conceived, so that

> However we reconstrue Stevens' law, it cannot be construed as one relating sensations to stimuli, since the former are incapable of measurement [p. 541].

More than half a thousand pages, it should be said, are devoted by Savage to the thesis that we cannot measure what psychophysicists seem so delightedly to be measuring.

Philosophical problems must not detain us, however, for philosophical issues are eternal and do not go away. More tractable, perhaps, are some of the specific issues that strike directly at the nomothetic imperative. A cluster of questions has concerned the problem of picking the best sensory scale.

When three different scales have been created on the same continuum by three separate sets of operations—jnd's, category estimation, and magnitude matching—on what grounds do we presume to ascribe to one of those scales a superior position? Helson (1964) seemed to deny that we can make any such judgment, for he said that

> no one scale, however carefully established, can be considered better than other scales obtained under different conditions of judging [p. 179].

Other authors have also lamented the absence of criteria for determining the validity of a choice of one kind of scale over another. Validity is indeed the issue here. Which scale best measures what it is that we want to measure? Since it is that kind of question, the answer becomes a matter of opinion—a value judgment. It appears that all problems related to validity must seek their ultimate solution in the pragmatic domain. If a scale does the job we want done, we usually accept it.

But opinions differ about problems and solutions. How else can we understand the decision of an experimenter to limit the observer's responses to a finite set of numbers, such as 1 to 7 or 1 to 20? That curious maneuver of constraining the observer's responses is a tactic that seems somewhat mulish to those who have allowed the observer a full range of responses and have witnessed the greater usefulness of the resulting ratio scales. It may be true that in the long run, superior procedures tend to replace inferior procedures, but in almost two decades of practice, magnitude estimation has not displaced category estimation—nor does it seem likely to do so any time soon.

Even among those who have given up category scaling in favor of ratio or magnitude matching, there remains a matter of taste and opinion that divides the practitioners. One view holds that the construct we call sensory magnitude follows simple laws, and in particular that under proper circumstances, the sensation magnitude experienced by the typical (median) observer grows as a power function of the stimulus magnitude. The power law is thought to set constraints on our expectations regarding the outcome of experiments, so that a discordant result becomes suspect until verified by adequate replication. In other words, the sensory power law takes precedence over the results of any particular experiment, just as the power law governing gravitational attraction usually remains unquestioned despite a particular experimenter's inability to confirm it by dropping objects in a laboratory. Both kinds of power laws make possible many kinds of predictions; yet they both can be shown to fail to some degree in particular contextual circumstances.

The other view holds that the nomothetic imperative has no compelling jurisdiction in psychophysics, because departures from the power law are too numerous to be ignored. It is claimed that the ways in which the human observer responds to stimulus intensity depend on prior learning, adaptation, range of stimuli, nature of the matching task, and so on; and that until those many factors and contextual influences can be discovered, explored, and understood, it is premature to speak of a psychophysical law.

The two views sketched above may prove more extreme than the attitude of any particular scientist, but they illustrate two poles of opinion. Perhaps the least nomothetic view yet expressed was that of Poulton (1968) who concluded his review of "the new psychophysics" by saying, "The mechanism of response learning and response bias must be included in any adequate description. To this reviewer," he added, "they present the more interesting and challenging problems." In my own view, the problems of response bias rate no better than a nuisance, an interesting nuisance, perhaps, as some of the foregoing sections have tried to show, but nevertheless a diversion from the basic business of sorting out the fundamental principles. Fortunately for science, however, its practitioners are motivated by a diversity of values and interests. It would stifle the enterprise if we all tried to crowd in on one single problem.

REFERENCES

BAIRD, J. C. *Psychophysical analysis of visual space.* Oxford: Pergamon Press, 1970.

BECK, J., & SHAW, W. A. Ratio-estimations of loudness-intervals. *American Journal of Psychology,* 1967, **80**, 59–65.

BERGLUND, B., BERGLUND, U., ENGEN, T., & EKMAN, G. Individual psychophysical functions for twenty-eight odorants. *Perception and Psychophysics,* 1971, **9**, 379–384.

DAWSON, W. E. An experimental analysis of judgments of sensory difference. Unpublished doctoral dissertation, Harvard University, 1968.

EKMAN, G. Two generalized ratio scaling methods. *Journal of Psychology,* 1958, **45**, 287–295.

FAGOT, R. A. On the psychophysical law and estimation procedures in psychophysical scaling. *Psychometrika,* 1963, **28**, 145–160.

GARNER, W. R. A technique and a scale for loudness measurement *Journal of the Acoustical Society of America,* 1954, **26**, 73–88.

GEIGER, P. H., & FIRESTONE, F. A. The estimation of fractional loudness. *Journal of the Acoustical Society of America,* 1933, **5**, 25–30.

HAM, L. B., & PARKINSON, J. S. Loudness and intensity relations. *Journal of the Acoustical Society of America,* 1932, **3**, 511–534.

HELLMAN, R. P., & ZWISLOCKI, J. J. Loudness determinations at low sound frequencies. *Journal of the Acoustical Society of America,* 1968, **43**, 60–64.

HELSON, H. *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row, 1964.

INDOW, T., & STEVENS, S. S. Scaling of saturation and hue. *Perception and Psychophysics,* 1966, **1**, 253–272.

JONES, F. N., & MARCUS, M. J. The subject effects in judgments of subjective magnitude. *Journal of Experimental Psychology,* 1961, **61**, 40–44.

KRANTZ, D. H. *A theory of magnitude estimation and cross-modality matching.* (MMPP 70-6) Ann Arbor: Michigan Mathematical Psychology Program, July 1970.

KRUEGER, L. E. Apparent combined length of two-line and four-line sets. *Perception and Psychophysics,* 1970, **8**, 210–214.

MARKS, L. E. Stimulus-range, number of categories, and form of the category-scale. *American Journal of Psychology,* 1968, **81**, 467–479.

MERKEL, J. Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studiën,* 1888, **4**, 541–594.

MILLER, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America,* 1947, **19**, 609–619.

MOSKOWITZ, H. R. Scales of intensity for single and compound tastes. Unpublished doctoral dissertation, Harvard University, 1969.

NEWHALL, S. M. A method of evaluating the spacing of visual scales. *American Journal of Psychology,* 1950, **63**, 221–228.

PLATEAU, J. A. F. Sur le mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Academie Royale de Belgique,* 1872, **33**(Ser. 2), 376–388.

POLLACK, I. Iterative techniques for unbiased rating scales. *Quarterly Journal of Experimental Psychology,* 1965, **17**, 139–148.

POULTON, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin,* 1968, **69**, 1–19.

POULTON, E. C. Choice of first variable for single and repeated multiple estimates of loudness. *Journal of Experimental Psychology,* 1969, **80**, 249–253.

SAVAGE, C. W. *The measurement of sensation.* Berkeley: University of California Press, 1970.

STEVENS, J. C., & GUIRAO, M. Individual loudness functions. *Journal of the Acoustical Society of America,* 1964, **36**, 2210–2213.

STEVENS, J. C., MACK, J. D., & STEVENS, S. S. Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology,* 1960, **59**, 60–67.

STEVENS, J. C., & MARKS, L. E. Spatial summation and the dynamics of warmth sensation. *Perception and Psychophysics,* 1971, **9**, 391–398.

STEVENS, J. C., & STEVENS, S. S. Brightness function: Effects of adaptation. *Journal of the Optical Society of America,* 1963, **53**, 375–385.

STEVENS, J. C., & TULVING, E. Estimations of loudness by a group of untrained observers. *American Journal of Psychology,* 1957, **70**, 600–605.

STEVENS, S. S. On the theory of scales of measurement. *Science,* 1946, **103**, 677–680.

STEVENS, S. S. On the brightness of lights and the loudness of sounds. *Science,* 1953, **118**, 576. (Abstract)

STEVENS, S. S. Biological transducers. *Convention Record, Institute of Radio Engineers,* 1954, Part 9, 27–33.

STEVENS, S. S. The measurement of loudness. *Journal of the Acoustical Society of America,* 1955, **27**, 815–820. (a)

STEVENS, S. S. On the averaging of data. *Science,* 1955, **121**, 113–116. (b)

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology,* 1956, **69**, 1–25.

STEVENS, S. S. Concerning the form of the loudness function. *Journal of the Acoustical Society of America,* 1957, **29**, 603–606. (a)

STEVENS, S. S. On the psychophysical law. *Psychological Review,* 1957, **64**, 153–181 (b)

STEVENS, S. S. On the new psychophysics. *Scandinavian Journal of Psychology,* 1960, **1**, 27–35.

STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* Cambridge, Mass.: M.I.T. Press, 1961. (a)

STEVENS, S. S. To honor Fechner and repeal his law. *Science,* 1961, **133**, 80–86. (b)

STEVENS, S. S. Power-group transformations under glare, masking, and recruitment. *Journal of the Acoustical Society of America,* 1966, **39**, 725–735.

STEVENS, S. S. Tactile vibration: Change of exponent with frequency. *Perception and Psychophysics,* 1968, **3**, 223–228. (a)

STEVENS, S. S. Measurement, statistics, and the schemapiric view. *Science,* 1968, **161**, 849–856. (b)

STEVENS, S. S. On predicting exponents for cross-modality matches. *Perception and Psychophysics,* 1969, **6**, 251–256.

STEVENS, S. S. Neural events and the psychophysical law. *Science,* 1970, **170**, 1043–1050.

STEVENS, S. S. Sensory power functions and neural events. In W. R. Loewenstein (Ed.), *Handbook of sensory physiology.* Vol. 1. New York: Springer-Verlag, 1971.

STEVENS, S. S. Perceived level of noise by Mark VII and dB(E). *Journal of the Acoustical Society of America,* in press.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology,* 1957 **54**, 377–411.

STEVENS, S. S., & GREENBAUM, H. B. Regression effect in psychophysical judgment. *Perception and Psychophysics,* 1966, **1**, 439–446.

STEVENS, S. S., & GUIRAO, M. Loudness, reciprocality, and partition scales. *Journal of the Acoustical Society of America,* 1962, **34**, 1466–1471.

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

STEVENS, S. S., & GUIRAO, M. Loudness functions under inhibition. *Perception and Psychophysics,* 1967, **2**, 459–465.

STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *Journal of Experimental Psychology,* 1956, **51**, 71–78.

STEVENS, S. S., & STEVENS, J. C. *The dynamics of visual brightness.* Cambridge: Harvard University, Laboratory of Psychophysics, 1960.

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

TORGERSON, W. S. Distances and ratios in psychophysical scaling. *Acta Psychologica,* 1961, **19**, 201–205.

(Received January 8, 1971; revision received May 6, 1971)