A Metric for the Social Consensus

Methods of sensory psychophysics have been used to gauge the intensity of opinions and attitudes.

S. S. Stevens

Experiments in a dozen laboratories have shown how procedures developed for the scaling of sensory attributes such as brightness and loudness can measure human reactions to many forms of nonmetric stimuli. The procedure called magnitude estimation, for example, has been used to gauge the consensus concerning intensity or degree for such variables as strength of expressed attitudes, pleasantness of musical selections, seriousness of crimes, and other subjective dimensions for which the stimuli can be arrayed only on nonmetric or nominal scales. These applications of direct scaling presage an advance beyond the indirect methods developed by Thurstone (*1*), who first brought the logic of Fechner's psychophysics into the domain of attitudes and opinions. The take-off point for Thurstone's indirect procedures was the observed variability of human judgments. By way of one or another assumption, the aim was to proceed from units of variability to units of magnitude. The hope was to achieve a scale of equal intervals having an arbitrary zero, like the ordinary scale of temperature.

Evidence that the new procedures of direct scaling can produce scales of a higher order than those produced by the Fechner-Thurstone approach stems mainly from a direct comparison of the results of the two kinds of methods. If, as many have believed, progress in the behavioral sciences demands the creation of tools to quantify such elusive variables as opinions, attitudes, preferences, esthetic values, utility, and so forth, the outcome of more than a dozen different experiments lays a foundation for optimism. On many attitudinal continua, the Thurstonian scale generated by the unitizing of variability or confusion shows an invariant relation to the scale of magnitude, as determined by the newer and more direct procedures. The relation bears a remarkable resemblance to that found when the same two kinds of procedures are applied to judgments of continua whose stimuli are measurable on ratio scales —as, for example, when judgment is made of the apparent brightness of lights of different intensity, or the apparent length of lines of different physical length.

Direct estimation shows that perceived or apparent length is approximately proportional to physical length (*2*). Obvious as this result may appear, the assurance that a line 20 centimeters long is judged to look about twice as long as one 10 centimeters long constitutes an important check on the method of direct magnitude estimation. Moreover, the scale of apparent length provides one more instance of the psychophysical law, which states that equal stimulus ratios produce equal perceptual ratios (*3*). The Fechnerian scale of length, which we obtain by counting off units of variability, often called just-noticeable differences (JND), has a different form. The Fechnerian JND scale is approximately a logarithmic function of physical length, and hence of apparent length. The reason the Fechnerian scale is logarithmic is that, since error tends to be relative, the variability of judgment—the just-noticeable difference—is roughly proportional to the magnitude of the stimulus. A simple test of these relations is to plot the Fechnerian JND scale against the logarithm of the magnitude-estimation scale. The result usually approximates a straight line. Similarly, a straight line is obtained when the "pair-comparison" scale of Thurstone is plotted against the logarithm of the magnitude scale produced by direct estimations of nonmetric stimuli, such as, for example, the seriousness of criminal offenses. The constancy of these relations, regardless of whether the stimuli are measurable on physical scales or whether the stimuli have no known metric, suggests a degree of invariance in the field of psychological measurement that may open bright prospects.

Progress in Psychophysics

Since the 1930's, when a commercial need arose for a scale of subjective loudness (because the decibel scale did not sound like a good loudness scale), several methods have been developed to assess subjective magnitude (*4*). It has been learned that observers can match numbers to stimuli and stimuli to numbers; they can estimate the apparent ratios between stimuli, and they can adjust stimuli to produce prescribed apparent ratios. All these methods give results that are related to the stimulus values by a power function. The power function is a necessary consequence of the ratio invariance of the psychophysical law. If equal stimulus ratios produce equal perceptual ratios, the perceived magnitude ψ grows as the physical value φ raised to a power β

$$\psi = k\varphi^{\beta}$$

The measure of φ begins at threshold; k is a constant that depends on the units used. Each modality or continuum appears to have its characteristic exponent, ranging in value from 0.33 for the brightness of luminous fields to about 3.5 for the apparent intensity of electric current passed through the fingers. A convenient feature of the power function is that, in log-log coordinates, it takes the form of a straight line. The slope of the line gives the exponent of the power function.

Perhaps the most reassuring development in psychophysics is the validation of the various exponents by the procedure of direct cross-modality matching. Just as lights of different hue may be matched for brightness (as in heterochromatic photometry), so sensations in one modality may be matched to those in another. Thus a person may adjust the loudness of a sound in his ear to equal the apparent strength of 60-cycle-per-second vibration on his finger. When the vibration

The author is professor of psychophysics and director of the Laboratory of Psychophysics at Harvard University, Cambridge, Massachusetts.

is changed, the observer changes the loudness to match the new apparent strength. The example in the log-log coordinates of Fig. 1 shows how the matching function turns out to be a straight-line power function whose slope (exponent) is given by the ratio between the exponent for loudness (0.6) and the exponent for vibration (0.95).

The basis for this result is as follows. If, given an appropriate choice of units, two modalities are governed by the equations

$$\psi_1 = \varphi_1^{\alpha}$$

and

$$\psi_2 = \varphi_2^{\beta}$$

and if the subjective values ψ_1 and ψ_2 are equated by cross-modality matches at various levels, then the resulting equal-sensation function will determine a relation between the two kinds of stimuli of the form

$$\varphi_1^{\alpha} = \varphi_2^{\beta}.$$

In terms of logarithms,

$$\log \varphi_1 = (\beta/\alpha)(\log \varphi_2).$$

In other words, in log-log coordinates the equal-sensation function becomes a straight line whose slope is given by the ratio of the two exponents. An interconnected net of exponent values has been validated by this direct matching procedure (*5*). On more than a dozen different continua the power-function relation has been confirmed and the value of the exponent has been checked by cross-modality comparisons.

This rather happy, if unexpected, development has demonstrated an important possibility. The stimulus-response relations for all the sense modalities can be mapped out without resort to numerical estimation on the part of the observers. The power functions obtainable by cross-modality matching make methods like magnitude estimation entirely dispensable.

On the other hand, magnitude estimation has the great advantage of convenience. Its prescription is simple. It calls for presentation of a series of stimuli in irregular order, if possible in a different order to each observer. The instructions may be modeled on the following example.

You will be presented with a series of stimuli in irregular order. Your task is to tell how intense 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 intense, assign a number 20 times as large as the first. If it seems one-fifth as intense, assign a number one-fifth as large, and so forth. Use fractions, whole numbers, or decimals, but make each assignment proportional to the intensity as you perceive it.

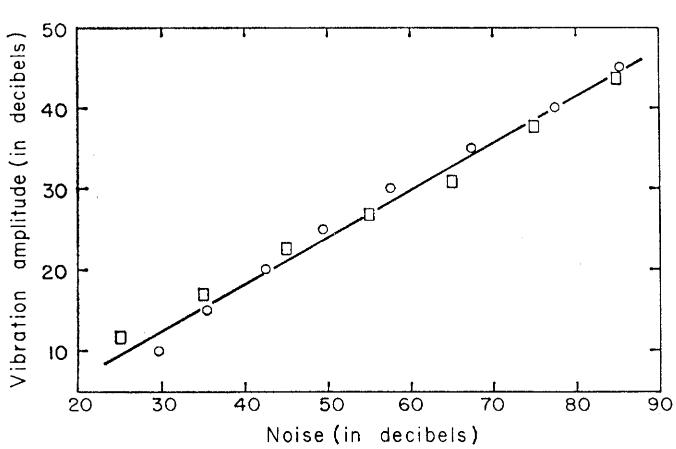
Because not everyone is familiar with the concept of proportionality, it has sometimes proved helpful to start off with an experiment on apparent length of lines. The lines, six to ten in number, should cover a wide range of lengths—say, a ratio of about 50 to 1. After judging such lines in irregular order, most observers seem to achieve a reasonably firm grasp on the concept of assigning numbers proportional to magnitude.

The variability of magnitude estimations has been found to grow approximately in proportion to the magnitude, and to produce distributions that are roughly log normal. Consequently, averaging is done best by taking geometric means of the estimations. This method of averaging also has the advantage that, despite the different ranges of numbers used by different observers, no normalizing is needed prior to averaging.

Three Kinds of Measures

In the course of the development of these new facets of psychophysics, it has been possible to clarify the relations among the three principal types of measures commonly encountered.

Magnitude measures. As already noted, the magnitude of a sensation is a power function of the stimulus. This is true for all prothetic (intensive) continua thus far explored—a total of about three dozen. The power function may or may not obtain on metathetic (qualitative) continua, such as the pitch of a sound or the position of a point.



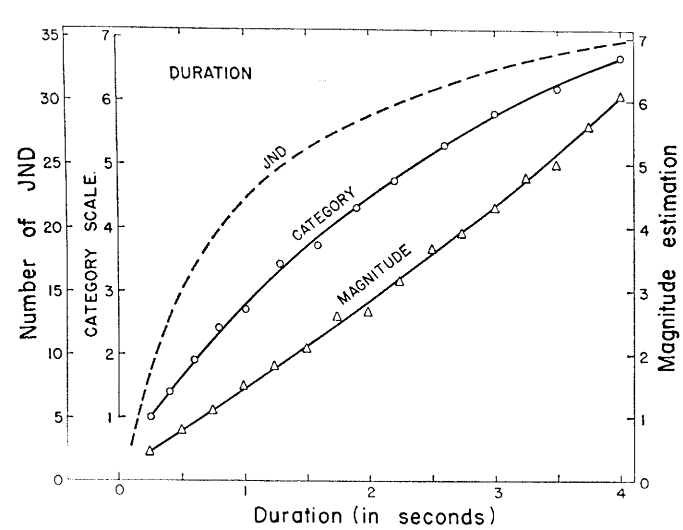


Fig. 1 (left). Equal-sensation function for cross-modality matching between loudness and vibration. The squares indicate that the observers adjusted the intensity of vibration on the fingertip to match the loudness of a noise delivered by earphones. The circles indicate that the observers adjusted the loudness to match the vibration. Each point is the decibel average of 20 matches, two by each of ten observers. Fig. 2 (right). Just-noticeable difference (JND) scale, category scale, and magnitude scale for apparent duration. (Triangles) Mean judgments of 12 observers who estimated the apparent duration of a white noise. The stimuli were presented in a different irregular order to each observer. (Circles) Mean category judgments made by 16 observers on a scale from 1 to 7. The end stimuli were presented at the outset to indicate the range, and each observer judged each duration twice in an irregular order. (Dashed curve) The function obtained by summating just-noticeable differences (JND).

Partition measures. All scales produced by asking observers to judge or produce what seem to be equal-appearing intervals belong to the general class called partition scales. Perhaps the most common form of this measure is the rating scale consisting of a prescribed set of categories to which stimuli are to be assigned, such as large, medium, and small. The result is commonly called a category scale.

When observers try to partition a continuum into equal intervals, the result, if successful, ought logically to represent a linear segment of the magnitude scale. Curiously enough, on prothetic continua it seldom if ever does. On prothetic continua the partition scale is practically always nonlinear relative to the magnitude scale. A typical result, one of dozens, is shown in Fig. 2. Relative to the magnitude scale, the curvature of the category scale is usually intermediate between linear and logarithmic.

On metathetic continua the partition scale tends to be linearly related to the magnitude scale. Thus the mel scale of pitch was constructed by combining the results of a partitioning experiment (equisection) with the results of a ratio-production experiment. The two procedures gave results that were in sufficient agreement to justify their combination (*6*).

Confusion measures. The measurement of a just-noticeable difference typically requires a person to categorize stimuli as greater or smaller than a standard. The stimuli are chosen to cover such a small range that the judgments are necessarily inconsistent. The problem is to find the stimulus difference that will evoke the judgment "greater" on 75 percent of the presentations. That is the difference usually employed to define the just-noticeable difference. As an aid in the computation process, the percentage of judgments called "greater" are often plotted against the stimulus values to form a poikilitic or scatter function, usually an S-shaped curve. These functions were called "psychometric" by Urban (*7*) in 1908 in the belief that the just-noticeable difference provided a unit, and hence a metric, for the measurement of psychological magnitudes. With the passing of that belief, it seems appropriate to call the functions by a more descriptive name. The term *poikilitic* suggests that the function displays the scatter, variability, or confusion that obtains among the observer's responses (*8*).

The summation of just-noticeable differences—the counting off of equivariability steps along the stimulus continuum—produces a Fechnerian scale, an example of which is shown in Fig. 2. A Fechnerian JND scale is a confusion scale in the sense that the basic operation for determining the steps along the scale is the assessment of equal degrees of confusion in the observer's responses.

Thurstone perceived the further possibility that, if confusion or dispersion can be used to create a scale in psychophysics, it can also be used to create scales in other contexts, for variability characterizes all human judgment. Thurstone's famous "equation of comparative judgment," together with its various elaborations, provides a means for transforming measures of variability, dispersion, or confusion into units of a scale. The confusions themselves may derive from any of a variety of procedures, among them pair comparisons and category scaling. With pair comparisons the observer says which stimulus in each pair of stimuli has more of some attribute. The inconsistencies among the choices provide the data for scaling. With category scaling, it is the confusions among the categories that are processed. The resulting scale goes by many names, but ought probably to be called the category confusion scale.

As noted above, when scales on prothetic continua are erected by direct magnitude estimation and are compared with scales derived by the procedures of Fechner and Thurstone, the relation is found to be approximately logarithmic (*9*). This relation provides a test that can be applied to nonmetric continua, such as seriousness of crimes, in order to determine the nature of the continuum. If the same relations that have been shown to obtain in sensory psychophysics among the three general kinds of measures can also be shown to characterize the comparable scales created with nonmetric stimuli, added confidence attaches to the outcome.

Attitude Scale

In 1929 Thurstone and Chave (*10*) used a version of category scaling in order to scale the strength of the attitude expressed by each of 130 statements concerning one or another aspect of religion. The resulting scale, based on the responses of 300 subjects, achieved a well-deserved fame, for it represented a serious and effective effort to introduce a metric where none had existed before.

In 1959 Finnie and Luce (*11*) undertook to apply a larger battery of scaling devices to some of the same attitude statements. When they used Thurstone's method of sorting the statements into 11 categories, the resulting partition scale correlated highly with the original partition scale of Thurstone and Chave. The passage of 30 years and the use of a new sample of subjects apparently made little difference.

A large difference resulted, however, when the subjects made magnitude estimations of the strength of the attitudes expressed by the various statements and the results were compared with the category confusion scale for the same items. The relation then became highly nonlinear, as shown in Fig. 3. The curve of Fig. 3, relating the category confusion scale and the scale obtained by direct estimation, is roughly logarithmic. That is the approximate relation to be expected if the perception of apparent strength of attitude constitutes a prothetic continuum.

Finnie and Luce say of the direct scaling procedure: "In addition to the theoretical interest in extending those methods and relations to areas other than psychophysics, knowledge of such a relation [a relation such as that of Fig. 3] can have considerable practical benefit. A magnitude scale on 10 or 100 items can be obtained from a group of people in, literally, a matter of minutes. The corresponding data for a Thurstonian analysis . . . requires more time to collect and is considerably more expensive to analyze, even with modern computation aids."

The ease with which a magnitude scale can be obtained from subjects instructed to match numbers to assorted items in a manner that preserves a proportional relation is indeed impressive, but a serious scaling venture will usually demand additional studies and may profitably be validated by other procedures, including cross-modality matching. As in any empirical inquiry, much depends on the level of accuracy required.

Thurstone and Chave went on to use their attitude scale to measure the attitudes of individuals drawn from various classes of people—college freshmen, graduate students, divinity students, and so on. An individual's position on the attitude scale is determined by the statements he is willing to endorse. As expected, the divinity students endorsed only the statements that were favorable to the church. The attitudes of graduate students, on the other hand, showed a wide spread.

Ekman's Law

At about the same time that Finnie and Luce were scaling attitude statements at Harvard University, experiments were under way at the University of Stockholm designed to show the relation between the Thurstonian indirect methods of scaling and the direct procedures of ratio and magnitude estimation. On seven different nonmetric continua thus far studied in Stockholm, the Thurstonian confusion scale has been shown to be a logarithmic function of the magnitude scale. Since this effort, led by Gösta Ekman, is perhaps the most extensive yet undertaken, the case for the use of the direct scaling methods in all areas involving human judgment receives much of its support from the Stockholm studies.

Ekman (*12*) has also been at pains to establish the generality of a principle that was conjectured in 1874 by Brentano and that has been demonstrated by others (*13*). The principle states that variability measured in psychological units is linearly related to psychological magnitude measured in the same units. This relation is the analog of Weber's law, which concerns the linear growth of variability among human judgments as a function of the stimulus measure. Both Fechner (in deriving his law) and Thurstone (in his "case V") proposed a different assumption—namely, that variability in psychological units is constant along the psychological continuum. That assumption has been found to be adequate for metathetic continua, but not for prothetic continua. As Ekman and his collaborators (*14*) have repeatedly demonstrated, on prothetic continua the variability, in subjective units, tends to grow as a linear function of the subjective magnitude.

Since it seems more important to have established the empirical generality of a principle than merely to have conjectured it as Brentano did, it would appear that the principle of the linear growth of subjective variability deserves to be called Ekman's law, the subjective counterpart of Weber's law.

The importance of Ekman's law in the present context is clear. Although the principle was developed and tested with metric stimuli in the area of psychophysics, there is every reason to believe that the same principle applies when the stimuli are describable only in terms of a nominal scale. Indeed, the repeated finding that the Thurstonian scale for nonmetric stimuli is

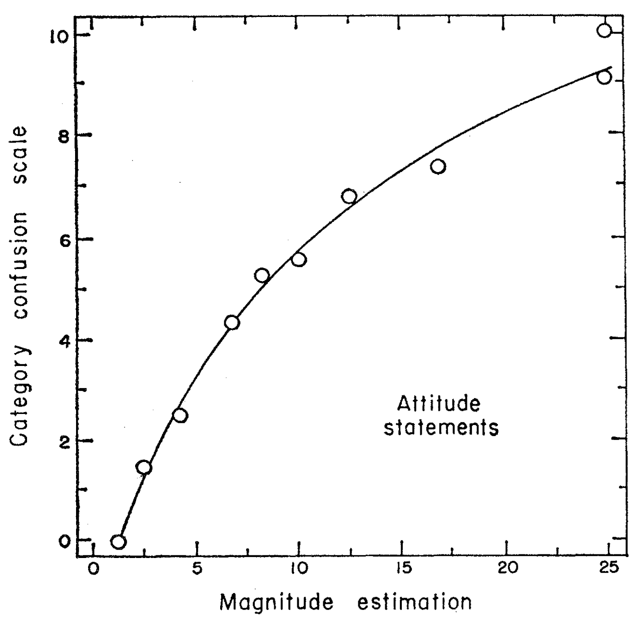


Fig. 3 (above). Comparison of two kinds of scales obtained from observers who judged the strength of the attitude expressed by various statements concerning the church.

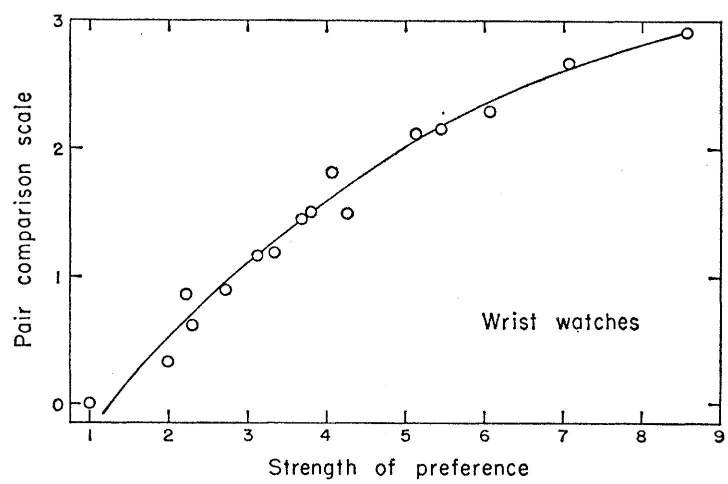


Fig. 4 (top right). Plot showing the curved relation between the pair-comparison scale (case V) and the scale of preference for different kinds of wrist watches, derived from ratio judgments made by adjusting line segments.

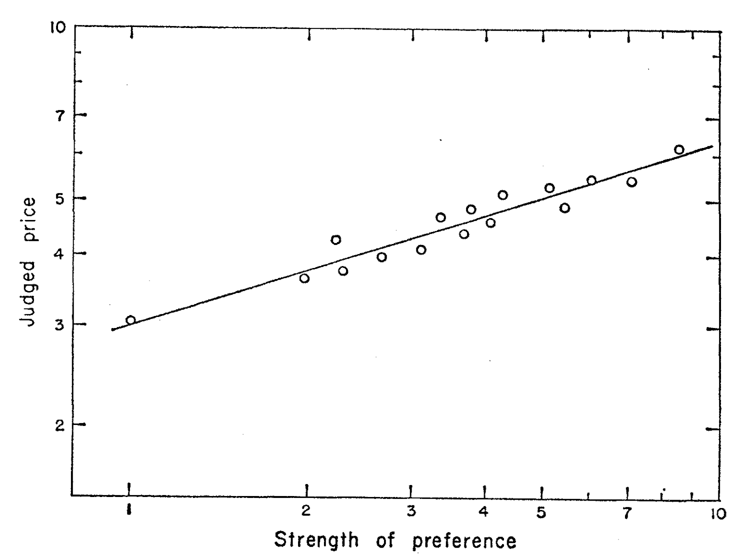


Fig. 5 (bottom right). The judged fair price (in yen $\times 10^3$) for various wrist watches plotted against the judged strength of preference. The coordinates are logarithmic. The line through the data points determines a power function with an exponent of 0.32.

related to the scale of magnitude in a logarithmic manner constitutes direct evidence in support of Ekman's law.

Thurstonian scales are usually erected on the assumption (case V) that subjective variability is constant all up and down the continuum. If Thurstone had gone further and had made the assumption that I have called case VI (*15*), he would, in effect, have been adopting Ekman's principle. Since case VI assumes a direct proportionality between variability and magnitude, under this assumption the units on the Thurstonian scale would mark off equal ratios rather than equal distances. An equal-ratio scale would be a power function of the magnitude scale, but the exponent would be arbitrary.

Preferences for Wrist Watches

Before turning to the Stockholm studies, we may note how busily the *Zeitgeist* has been at work producing similar experiments in different places. In the same year, 1959, Indow (*16*) was presenting pictures and descriptions of wrist watches to 127 university students in Japan and asking for pair comparisons and ratio judgments. The student was asked to say which watch in each pair of watches he preferred, and also to indicate what the relative strength of his preference was by marking a position on a line 8 centimeters long. The students in effect matched length of line to subjective value, a kind of cross-modality procedure. By comparing the ratio scale determined with the aid of the 8-centimeter line to the Thurstonian scale (case V) derived from the "noise" or confusions in the pair comparisons, Indow was able to demonstrate an approximately logarithmic relation between the two kinds of scales. This nonlinear relation is shown in Fig. 4.

In another part of Indow's experiment the students were asked to state what they would regard as a fair price for each of the wrist watches, if the price of a particular watch was a stated number of yen. For each of three different groups of students a different standard watch and price were designated. The interesting aspect of the outcome concerns the relation of the averages of the price estimates to the ratio scale of desirability previously established. This relation is shown in log-log coordinates in Fig. 5. The fact that the data approximate a straight line suggests that fair price is judged to be a power function of degree of desirability. The low value of the exponent, 0.32, indicates that the relation between judged price and judged desirability is rather far from linear. It would be interesting to know whether a power function with a similarly low exponent would be obtained in other circumstances.

Esthetic Value of Handwriting, Drawings, Music

Two separate studies of the esthetic value of handwriting were carried out by Ekman and Künnapas (*17*). In both studies, samples of handwriting were scaled by Thurstone's method of pair comparisons and by a variant of the method of ratio estimation (*18*). The first study, in which seven samples of handwriting were used, is especially instructive because it failed to show a logarithmic relation between the scale by pair comparisons and the scale by ratio estimation. In the second study, 18 samples of handwriting were used, covering a wider range of quality. This second experiment demonstrated that the approximately linear relation of the first experiment became an obviously logarithmic relation when the wider stimulus range made the form of the function easier to determine.

A related study by similar methods was carried out on 17 drawings of a tree. The samples were selected from some 200 drawings produced by sixth-grade students (*19*). Again it was found that the confusion scale derived from pair comparisons was quite accurately proportional to the logarithm of the magnitude scale derived from ratio estimations.

Esthetic judgment in music was investigated by Koh (*20*), who presented 51 vocal selections and 60 piano pieces to various populations of subjects, including college students and patients in the alcoholic ward of a hospital in North Dakota. Each excerpt lasted about 15 seconds for the piano pieces and about 60 seconds for the vocal pieces. The total population of subjects, numbering 330, was divided into six groups, two groups each of college males, college females, and alcoholics. Half of the groups, one of each category, heard vocal selections, the other half heard piano selections. The subjects in each group made magnitude estimations of the affective value of each selection and also judged each selection on a category scale expressed in terms of nine adjectives ranging from "most pleasant," through "indifferent," to "most unpleasant." The category values were treated as a 9-point numerical scale, and the ratings for each piece of music were averaged.

For all six groups, the average rating scale value was approximately proportional to the logarithm of the geometric means of the magnitude estimations. The product-moment correlations between the category values and the logarithms of the magnitude values ranged from 0.90 to 0.96. For some groups there was a slight upward concavity of the curve; this is usually the case when category scales are plotted against the logarithm of the magnitude scale.

As Koh remarked, the relation between the category and the magnitude scales was strikingly invariant in spite of the differences in the age, sex, education, occupation, and pathology of the subjects. "These empirical invariances," he concluded, "strongly suggest the usefulness of magnitude estimation for complex judgmental processes."

Importance of Swedish Monarchs

The direct method of ratio estimation and the indirect method of pair comparisons were used by Ekman and Künnapas (*21*) to construct scales of the judged political importance of 11 Swedish monarchs who lived between 1550 and 1850. Eighty-three students of psychology made the judgments. The scale, derived from pair comparisons based on the assumption of Thurstone's case V, was a logarithmic function of the magnitude scale derived from direct ratio estimation. The scale of magnitude shows that, on the average, these students considered the leading monarch to be about four times as important as the monarch regarded as least important. Whether it is useful to determine ratios among items of this kind remains to be seen. The point to note here is that ratio determinations are possible.

Occupational Preference

Perloe measured the degree of prestige that attaches to each of 100 different occupations by means of two procedures: magnitude estimation and a 7-point category rating scale (*22*). The subjects were 40 undergraduates at Haverford College. Despite a certain mix-up about the instructions, the data show that the relation between the category scale and the magnitude scale is essentially the same for judgments of occupations as it was for judgments of musical selections in Koh's experiments. As with loudness, brightness, and other attributes for which there exists a stimulus metric, the mean category judgments define a scale that is almost, but not quite, a logarithmic function of the median magnitude estimations.

This is the expected outcome when the "noise" or variability in the experiment is large. When the variability is small, the category scale departs farther from the logarithmic form (*23*). Under more favorable circumstances—where, for example, the subject may be permitted to adjust a stimulus to bisect the apparent distance between two other stimuli—the partition scale may approach fairly close to the magnitude scale (*24*).

In the judgments of those of Perloe's observers who appear to have grasped the instructions, the range from the most to the least prestigious occupation was about 30-fold.

A roughly comparable range characterized the judgments of 74 students at the University of Stockholm who expressed their preference for 17 different occupations (*25*). In the Stockholm study two different procedures—ratio estimation and magnitude estimation—were used to scale occupational preference, and a third procedure—pair comparisons processed according to the assumption of case V—was used to produce a Thurstonian scale. The two magnitude scales were found to be linearly related, as had been expected. The confusion scale derived from pair comparisons approximated a logarithmic function of both magnitude scales.

The 17 occupations are shown in Fig. 6 in the positions assigned them by the geometric means of the magnitude estimations. Among university students in Sweden, the occupation of physician appears to be rated far out in front.

Pleasantness of Odors

The scaling of odors presents an especially instructive example of the potentialities of the ratio-scaling methods. The apparent intensity of each of a variety of odors has been scaled by magnitude estimation, always with approximately the same results. The apparent intensity of the odor grows according to a power function of the concentration, and the exponent is of the order of 0.5 or 0.6. Furthermore, the range from the faintest to the strongest perceptible odor is rather small—nothing like the wide dynamic range of loudness or brightness.

The subjective range increases to a ratio of more than 100 to 1 when, instead of the intensity of a single odor, the pleasantness of 18 different odors is the attribute that is estimated (*26*). The hedonic aspect of odor exhibits a wider dynamic range than the intensity aspect.

When the pleasantness of the 18 odors was judged on a nine-point category scale, the result was typical of category judgments on other prothetic continua. Plotted against the scale produced by magnitude estimation, the category scale of pleasantness gives a curve that is concave downward. Plotted against the logarithm of the magnitude scale, the category scale

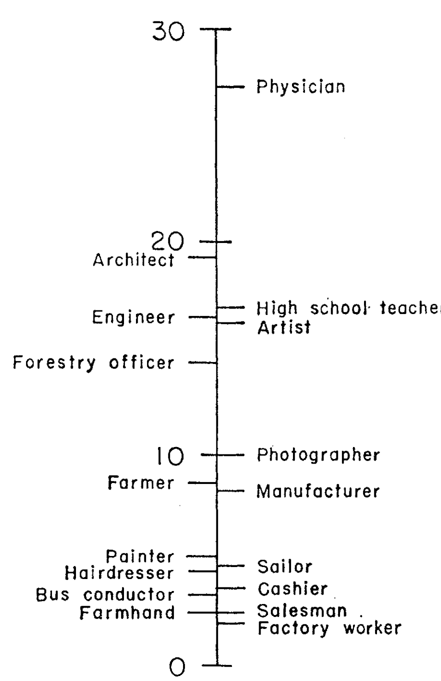


Fig. 6. Degree of preference for various occupations expressed by students at the University of Stockholm.

gives a curve that is concave upward. This latter relation demonstrates the principle that the curvature of the category scale is generally intermediate between a linear and a logarithmic function of the magnitude scale.

Liberal-Conservative

Thus far, there appears to be one possible exception to the finding that a scale based on the unitizing of variability approximates a logarithmic function of the magnitude scale generated by the direct estimation of some attribute of a series of nonmetric stimuli. Ekman and Künnapas (*27*) assembled a set of 17 statements designed to sample the liberal-conservative continuum and asked 82 subjects to assess the degree of conservatism expressed by each statement. As it turned out, the scale defined by ratio estimations was roughly linearly related to the scale defined by the procedure of pair comparisons (case V). This unusual result was further examined in a series of experiments by Künnapas and Sillén (*28*), who used as stimuli the same 17 statements used in the first study. In this second study, judgments were also made of the degree of liberalism expressed by the statements. Except in one part of the second study, there was no clear evidence for a logarithmic relation between the Thurstonian and the magnitude scales.

Künnapas and his associates entertain the hypothesis that this continuum is qualitative or metathetic; this, if true, would account for the general nature of the results. An alternative hypothesis is that the heterogeneity of content among the statements made the assessment of the degree or intensity of one particular attribute rather difficult. The statements concerned such diverse matters as world government, church and state, sex education, movie censors, school prayers, abortions, and capital punishment.

The problem is analogous to the task of judging loudness when every tone presented has a different pitch. Such loudness judgments have indeed proved possible, but they have also turned out to be more variable than judgments obtained when only the intensity of the tones is varied (*29*). It is possible that the 17 statements

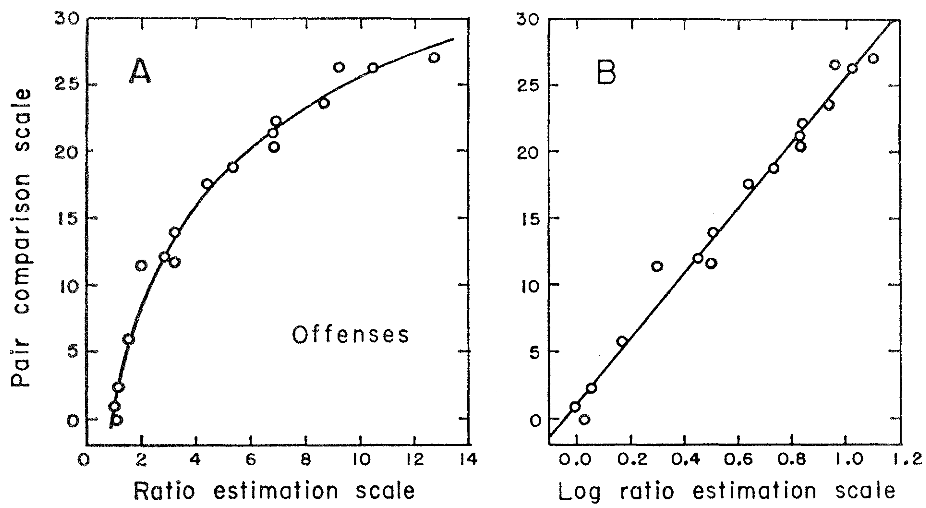


Fig. 7. Plots showing the relation between the confusion scale based on pair comparisons and the magnitude scale based on ratio estimations of the seriousness of offenses. (*A*) The relation in linear coordinates; (*B*) the relation in semilogarithmic coordinates. The variability or confusion among the pair comparisons grows as a linear function of the judged magnitude, which entails a logarithmic relation when units of variability are plotted against magnitude.

used in the experiments of Künnapas and his associates did not elicit judgments related to the same subjective continuum for all the subjects. By contrast, all the statements in the study by Finnie and Luce contained explicit reference to the church. Perhaps the chief problem in these areas is to refine the statements until a sufficient homogeneity is achieved to make it easy for the subjects to judge degree or intensity. Thurstone and Chave gave many helpful rules for the elimination of ambiguous and irrelevant statements and for the refinement of scale items.

With simpler materials, such as single adjectives describing personality traits, it is apparently quite possible for people to abstract an attribute and judge its intensity. Ekman and Künnapas (*30*) asked 95 subjects to judge the degree of masculinity expressed by each of 11 adjectives descriptive of personality. The subjects also judged the degree of femininity expressed. In both instances, the scale derived from pair comparisons (case V) approximated a logarithmic function of the scale derived from direct ratio estimation.

Seriousness of Offenses

On the basis of a preliminary study of opinion concerning a number of offenses, Ekman (*31*) selected descriptions of 17 more or less immoral actions for a study of moral judgment. The actions ranged from hit-and-run by a drunken driver down to stopping in a no-parking zone to mail a letter. Eighty subjects made pair comparisons and ratio estimations. As shown in Fig. 7, the pair-comparison scale based on a processing of the noise or confusion (case V) is very close to a logarithmic function of the scale based on direct ratio estimations.

Ekman's study, like most of those described thus far, was methodological in intent: the object of interest was method, not the achievement of a practical, substantive outcome. A full-length study in which method was the means rather than the end has been reported by Sellin and Wolfgang (*32*). Their 423-page book is directed at improvement of the methods used to compile police and court statistics for the purpose of measuring criminality in general and delinquency in particular. The research design placed major emphasis on delinquency events, not on delinquent persons, for the purpose was to measure the amount and type of harm to the community attributable to antisocial acts.

The general strategy of this 3-year study was as follows. First, a representative 10-percent sample of delinquency events was selected by random sampling from the universe of all such events in Philadelphia, Pennsylvania, recorded in the year 1960. Scaling procedures were then applied to events selected from the sample in order to convert the judged seriousness of the events into numerical scores. A final combination of all the information produced a delinquency index, a device that can be used to gauge the total incidence of delinquency and the effectiveness of whatever preventive measures may be brought to bear on the grave problem of antisocial behavior.

It is the second stage of the study that most concerns us here, the quantification of the gravity of delinquent acts, a quantification that must rest ultimately on the judgment of members of society. In brief outline, Sellin and Wolfgang proceeded as follows: A list of 141 offenses was first compiled, and a carefully phrased statement was made of each offense. These statements, typed on cards, were submitted for trial testing to 17 raters, mostly college students, who rated the seriousness of each offense on a 7-point category scale. Three representative offenses were then selected from each of the seven categories for use in further testing. These 21 offenses, presented in carefully randomized orders, were judged by 569 people—38 juvenile-court judges, 286 police officers, and 245 students from two universities. About half of each class of raters made magnitude estimations of the seriousness of the offenses; the other half rated the offenses on an 11-point category scale.

The next question of interest concerns the relation between the two kinds of judgments, category and magnitude. Figure 8 shows a direct comparison between the results for the 38 juvenile-court judges, 20 of whom used the category scale and 18 of whom made direct magnitude estimations. As is characteristic of prothetic continua, when the category scale of degree of delinquency is plotted against the magnitude scale, the curve is concave downward. When the same category ratings are plotted against the logarithm of the magnitude scale, the result is more nearly linear, but slightly concave upward, as shown in Fig. 9. Finally, when, instead of the averages of the category assignments, only the variability or confusion among the category assignments is used to generate a category confusion scale, the result is more nearly a linear function of the logarithm of the magnitude estimations, as shown in Fig. 10.

For different groups of raters, the age of the offender was specified as 13, 17, or 27 years, or it was left unspecified. Ten different plots like the plot of Fig. 10 were made from the judgments of the ten subgroups of raters.

The impressive feature of the ten plots is their invariant form. When the total ordinate scale is taken as one unit, the slopes of the ten functions range from 0.22 to 0.31. The slope in Fig. 10 is 0.29. No significant differences were attributable to the age of the offender. It was the offense itself that seemed to determine the judgment of seriousness.

More important, perhaps, there was also impressive invariance across raters. Juvenile-court judges produced scales comparable to those produced by police officers and college students. It may be surprising that all three clssses of raters agreed, for example, that stealing and abandoning a car is only about one-tenth as serious as robbing a man of $5 and wounding him in the process. According to the consensus, this latter crime becomes about two-and-a-half times as serious if the victim dies. Out of these magnitude estimations, say Sellin and Wolfgang (*32*, p. 268), "a pervasive social agreement about what is serious and what is not appears to emerge, and this agreement transcends simple qualitative concordance; it extends to the estimated degree of seriousness of these offenses."

The next major step in the study was an item analysis designed to refine further the statements used to define the offenses. The revised statements were used in a retest with a new population, a group of 195 students from still another university. This final testing gave results that correlated highly with the earlier data and thereby provided added justification for the construction of an index of delinquency based on a representative ratio scale of seriousness of offense.

An important feature of the final index is its provision for the additivity of offenses, a feature justified to a large extent by the outcome of the magnitude estimations of seriousness. Thus, the stealing of $5 is given a rounded value of 1. Breaking into a building also has the value 1. Breaking in *and* stealing $5 has the value 2, because the magnitude estimation score for the seriousness of the combined act was approximately double the estimate for each act separately. As another example, forcible rape has the value 11—8 for the forced sex act, 2 for the intimidation of the victim, and 1 for the inflicting of minor injury. The extraction of the additive components of the complex delinquent acts was achieved, of course, through the process of analyzing the results of the magnitude estimations. It is doubtful that any of the raters would have been conscious of the underlying additivity in any explicit way, and some of them would probably be offended by the thought that one forcible rape can be equated to some number of money thefts. Nevertheless, both the quantitative estimates of large numbers of raters and the gradations in the punishments prescribed by law make a strong argument for equatability and additivity among offenses.

The Value of Money

Among the offenses rated by magnitude estimation there were a few that dealt with the stealing of money. Statements naming stolen amounts of 5, 20, 50, 1000, and 5000 dollars had been scattered at random among the other statements describing offenses. The geometric means of the judged magnitude of the seriousness of the offenses involving money grew as a very precise power function of the amount of

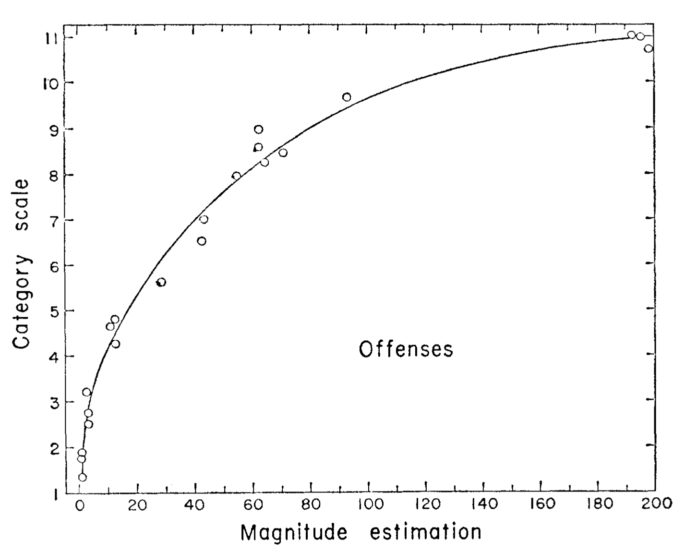


Fig. 8 (above). Relation between the means of the category judgments and the geometric means of the magnitude judgments for seriousness of offenses, as rated by juvenile court judges. The coordinates are linear.

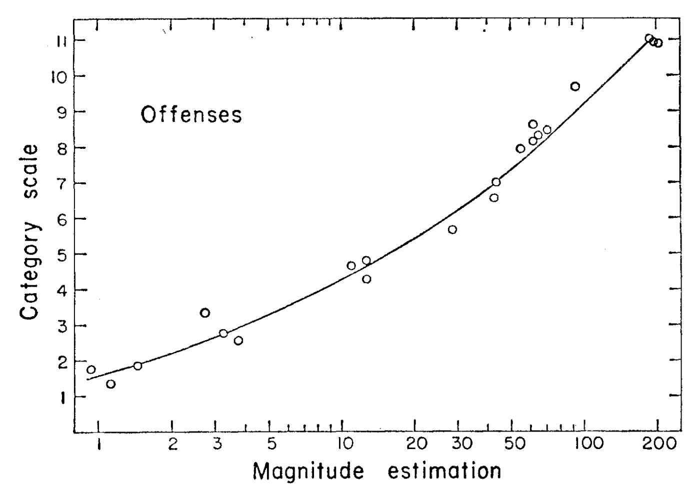


Fig. 9 (top right). Relation between the category scale and the magnitude scale for seriousness of offenses, as rated by juvenile court judges. The coordinates are semilogarithmic. The data are the same as in Fig. 8.

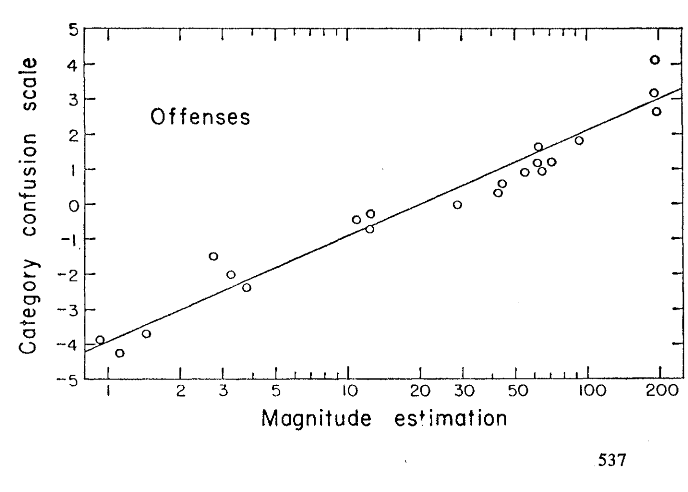


Fig. 10 (right). Relation between the category confusion scale and the magnitude scale, in semilogarithmic coordinates. The data are from Fig. 8. The ordinate represents the Thurstonian scale derived from the confusion or variability among the judgments made on the category scale of seriousness of offenses.

money stolen (Fig. 11). The value of the exponent, 0.17, suggests that, in order for one crime to be considered twice as serious as another, the amount stolen must be about 60 times as large. Although it is worse to steal $2 than $1, it is certainly not twice as bad.

The argument that the subjective value of money, the economists' "utility," may be a power function of the number of dollars goes back to the mathematician Gabriel Cramer (*33*), who made the conjecture in 1728. Ten years later, and quite independently, Daniel Bernoulli hypothesized a different law, a logarithmic function (*33*). Bernoulli's proposal became very well known, while Cramer's was long forgotten. But the power function, with an exponent smaller than 1, has empirical backing to recommend it (*34*). This same type of power function also accords with the universally agreed upon fact that to the average man each added dollar has less utility (value) than the dollar preceding in—the economic principle of "decreasing marginal utility." It is interesting that, in the context of loss due to acts of delinquency, the appraisal of the value of money also takes the form of a power function with an exponent of less than 1.

Punishment

How well does society's accumulated wisdom, or lack thereof, in legislating punishment accord with the judged gravity of offenses? In particular, what does the Pennsylvania Penal Code say about maximum penalties for the 21 offenses scaled in the main study? The answer is both interesting and encouraging. The product-moment correlation between seriousness of offense as judged by university students and maximum penalty stated in terms of time in jail was 0.88, provided a death sentence is interpreted as a jail term equal to the life expectancy of the median perpetrator of homicide. The correlation was even slightly higher, 0.94, for the magnitude-of-offense judgments by police officers. Both sets of results are shown in logarithmic coordinates in Fig. 12. As Sellin and Wolfgang express it (*32*, p. 327), "These correlations are surprisingly high considering the fact that the Penal Code provides no variation in the maximum [penalty] for amounts of money stolen and relatively few intervals between thirty days' imprisonment and death." Note also that the punishment scale is truncated at its lower end, for the smallest maximum penalty is 30 days in jail.

Another point of interest is the general form of the relation in Fig. 12. The straight line through the data represents a power function, because the coordinates are logarithmic. The slope of the line in these coordinates gives a measure of the value of the exponent. The slope (exponent) is clearly less than 1.0. Its value, 0.7, means that the penalty (time in jail) is not proportional to the seriousness of the offense. In linear coordinates the line in Fig. 12 would appear as a curve that is concave downward.

In order to have the full story on the justice of the maximum penalty specified by the Code, we would need to know another function: the judged seriousness or severity of various periods in jail—that is, the subjective value function for terms of imprisonment. Since that function was not directly scaled by Sellin and Wolfgang, we can approach the question indirectly by way of an assumption—the assumption that in this, the best of all possible worlds known thus far to *Homo sapiens*, the punishment fits the crime, provided both are assessed by direct subjective judgment. If that assumption holds, it follows that the subjective severity function for time in jail has an exponent that is the same as the exponent in Fig. 12—namely, 0.7. That exponent, with a value of less than 1, raises the interesting question whether people regard the severity of punishment that goes with various periods in jail as a decelerating function of calendar time. Is a sentence of 2 months less than twice as punishing as a sentence of 1 month? I would think so. If others agree, then perhaps the Pennsylvania Penal Code does indeed mete out roughly proportional justice.

In the preceding argument, the absolute value of the intercept of the hypothetical punishment function has been neglected. It will have to be considered, of course, before the complete story is told, because the *absolute* amount of punishment for a given offense merits as much concern as does the *relative* amount of punishment for different offenses.

However that may be, Sellin and Wolfgang have shown how to attack an urgent social problem with methods that were developed in psychophysics for the study of human sensory systems. The methods they borrowed have produced impressive and useful results. It is a large and onerous task to develop and refine, by repeated revisions, a scale with useful properties in an area as complex as delinquent behavior. The one-shot experiment, so typical of the academic investigator, will not suffice when the goal is serious and substantial. The ratio-scaling methods used by Sellin and Wolfgang have a long history, and it is instructive that their development got its biggest push from a practical problem in acoustics (*35*). In the 3-year study of delinquency, the extension of the ratio-scaling methods to social variables has been dramatically achieved, largely because the challenge of the problem has justified the investment needed to track down and eliminate needless sources of noise and variability. Science seems to do its best when it faces a problem worth solving.

Functions for Individual Observers

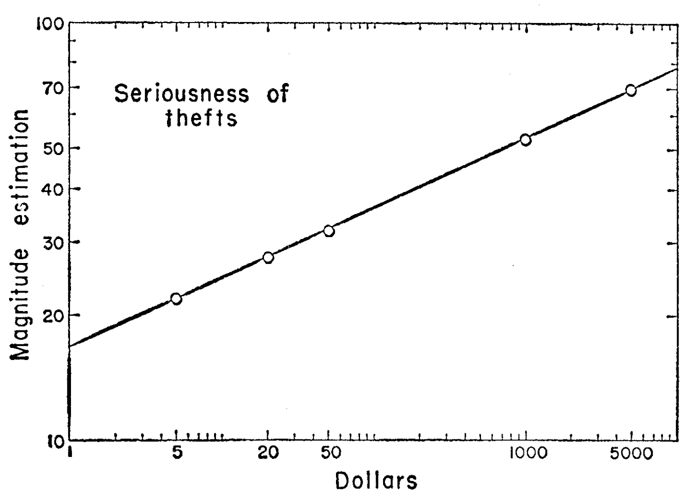
Assessments of subjective value like those shown in Fig. 12 represent averaged results, and it may be objected that the scale of seriousness of delinquent acts does not represent the opinion of some particular person. Indeed it may not, for the first task must be to discover the consensus, if there is one. After a representative value function has been spelled out, it may or may not become profitable to ask about the exceptions. Whether, in a given domain, the consensus is sufficiently homogeneous to justify averaging is an empirical question, and one that has begun to receive attention in psychophysics.

It was for groups of observers that loudness and brightness were first found, on the average, to grow in proportion to the stimulus intensity raised to a power (*36*). As might be expected, little attention was paid at first to the question whether the power law would describe the reactions of each individual. Understandably, therefore, the question arises whether the ubiquitous power function may not be the result of group averaging. Pradhan and Hoffman (*37*), upon finding that some of their six observers did not produce power functions when judging apparent weight, concluded that individual functions "were found not to follow Stevens' law although averaging over observers does yield a power function. Stevens' power function thus seems to be an artifact of grouping."

That negative note contrasts with the view expressed by Ekman and Sjöberg (*38*): "After a hundred years of almost general acceptance . . . , Fechner's logarithmic law was replaced by the power law. The amount of experimental work performed in the 1950's on this problem . . . was enormous. . . . The power law was verified again and again, in literally hundreds of experiments. As an experimental fact, the power law is established beyond any reasonable doubt, possibly more firmly established than anything else in psychology."

An empirical answer to the problem of the individual function calls for the straightforward procedure of measuring and exhibiting the functions for individual observers. An early attempt to exhibit a collection of individual power functions was vetoed by a journal editor who pointed out, quite rightly, that nothing was shown by 39 straight lines in log-log coordinates that could not be summarized in a sentence or two (*39*). Since then, the problem of the individual function has led members of the Laboratory of Psychophysics at Harvard to develop new procedures. Lee McMahon made the first experiments on loudness with a technique that leaves the observer free to set the level of the sound intensity and also to estimate the loudness. It is a combined production-estimation technique which produces data that cannot be averaged, because each observer sets different stimulus levels and makes different estimates, at his pleasure. He may be asked simply to set as many levels as he likes and to assign numbers proportional to the loudness as he hears it. Results obtained with this technique by J. C. Stevens and M. Guirao (*40*) are shown in Fig. 13.

If the power law can be verified in a sufficient number of experiments, per-



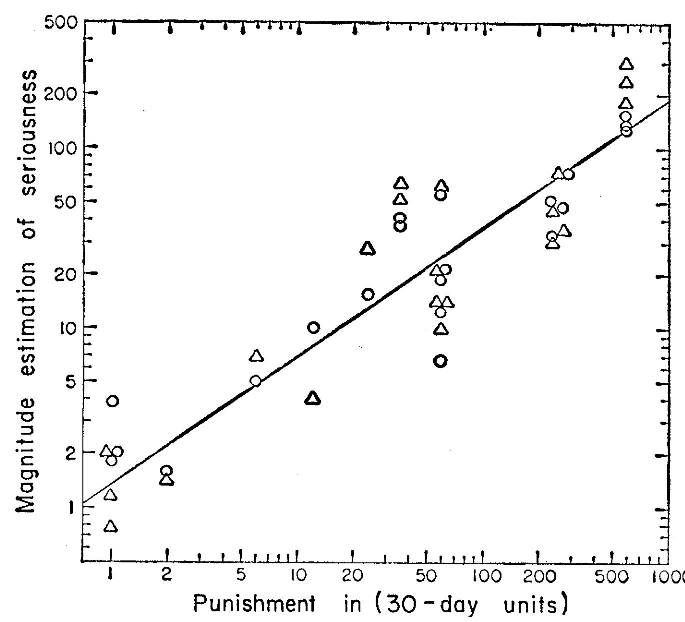
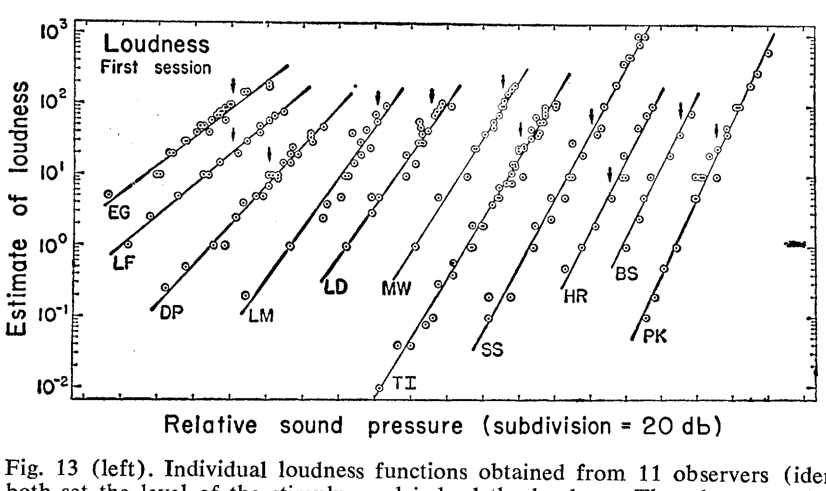


Fig. 11 (left). Magnitude estimations of the seriousness of stealing various amounts of money. The values on the ordinate are the geometric means of estimates made by 105 university students. The line, in log-log coordinates, defines a power function with an exponent of 0.17. Fig. 12 (right). Relation between the geometric means of the judged seriousness of 21 offenses and the maximum penalty prescribed by the Pennsylvania Penal Code. The raters were police officers (circles) and university students (triangles). For plotting purposes the police ratings were multiplied by 0.5. The line through the data has a slope of 0.7.



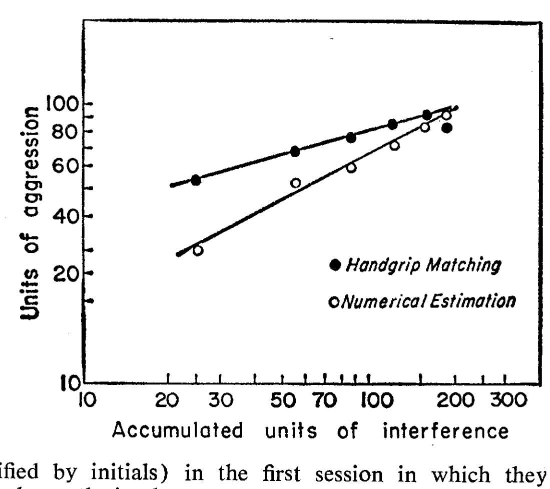


Fig. 13 (left). Individual loudness functions obtained from 11 observers (identified by initials) in the first session in which they both set the level of the stimulus and judged the loudness. The observers differed greatly in the range and number of stimuli they produced and estimated, but all the data approximate power functions. The small arrow above each function indicates a level of 80 decibels on the abscissa. Fig. 14 (right). Aggression, expressed as degree of dislike for a "leader" who interfered with the success of the subjects in a game, is plotted against the amount of the interference (number of points lost by reason of the leader's wrongheadedness). Forty subjects made magnitude estimations of their dislike and also squeezed a hand dynamometer by an amount proportional to the intensity of their feeling. The relative slopes of the two lines are predictable from other experiments.

haps the point of interest will shift to the opposite side: What characterizes an observer whose judgments do *not* give a power function when the data are cross-checked by a variety of procedures? A single experiment or a single technique may not be enough to show conclusively that, for a given person, the perceived magnitude fails to grow as a power function of the stimulus magnitude. Almost anything may happen in a single experiment. A firm decision about an individual case may call for a multiple attack by estimation procedures, production procedures, and cross-modality comparisons. Hopefully, a multiple approach may even prove adequate to reveal an abnormal sensory function, such as auditory recruitment, in individual patients.

Two features of Fig. 13 deserve comment. All 11 observers produced rather good approximations to power functions, but the slopes (exponents) varied from person to person. The analog of this second feature showed up also in the study of delinquency by Sellin and Wolfgang, where the range of the magnitude estimations for the seriousness of the offenses varied from person to person.

How should we regard these individual differences in range, or in exponent? Admittedly it would be something of a miracle if everybody's judgments followed exactly the same function for delinquency, loudness, or anything else. Perhaps the variations in how people use numbers and how they regard ratios are no more than the inevitable noise that characterizes these complex processes. The fact that two of the lines in Fig. 13 have different slopes may mean that the two observers in question have different mechanisms at work in their auditory systems, but it may also mean that the two observers happen merely to differ about what they consider an apparent ratio. Further experiments may decide the point. In the meantime, there is growing evidence that the differences in the observed exponents, for a reasonable sample of observers, have one of the very important properties of noise—namely, the capacity to be averaged out. Note, for example, in Fig. 12 how nearly the average estimations by the police officers agree with the average estimations by university students. It is the stability of the function from group to group that makes the result useful.

Scaling Frustration and Aggression

A review of the potential power and usefulness of psychophysical scaling procedures would be incomplete without mention of an attempt to produce measurable amounts of frustration in a group of subjects and to scale the consequent acts of aggression. A series of realistic and imaginative experiments were devised by Hamblin, Bridger, Day, and Yancey (*41*), in which genuine feelings of aggression were produced in college R.O.T.C. students by means of a simulated game.

The student could advise a "leader" concerning the tactics of the game, but the leader had the final say, and he could thereby cause game points to be lost. This interference with the goal of winning points may be taken as a measure of the amount of frustration. At various stages of the procedure the subjects were asked to estimate numerically the degree of their dislike for the leader, who had been represented to them as a candidate for promotion whose qualities of leadership they were to judge. They were also asked to squeeze a hand dynamometer to express the degree of their dislike for the leader. In expressing their dislike, the subjects were in effect showing aggression against the leader, because, to the best of their knowledge, their expressed dislike would injure his chance of promotion.

The results obtained from 40 subjects are plotted in log-log coordinates in Fig. 14. For both the magnitude estimation and the handgrip studies, the data approximate straight-line power functions. Hamblin and his associates suggest that the deviant point on the handgrip curve probably resulted from the fact that, for both the last value and the one before it, many subjects had squeezed "as hard as they could." However that may be, it is interesting to compare the slopes (exponents) of the two functions, 0.53 for magnitude estimation and 0.30 for handgrip. The ratio, 0.53/0.3 (=1.77), is close to the value found in psychophysical experiments when observers estimate directly the apparent force of their handgrips. In the experiment on frustration we seem to have another example of the utility of the cross-modality matching procedure—a procedure whose power and versatility are coming to be recognized.

Conclusions

What then are the invariances in these manifold experiments involving human judgment? A convergence of evidence from fields as disparate as psychophysics and criminology has pointed to stable and constant relations. One such relation states that subjective magnitude is a power function of stimulus magnitude. The underlying invariance then becomes the simple principle that equal stimulus ratios produce equal subjective ratios.

On many of the continua discussed above, the stimuli can be measured only on a nominal scale, for the stimuli are verbal statements, occupations, crimes, musical selections, and other nonmetric items. On those continua the power law cannot be confirmed directly, but there emerges another notable invariance.

For both kinds of continua, those based on metric stimuli and those based on nonmetric stimuli, there is a constant relation between the scale erected by direct judgment and the scale derived from a unitizing of variability or confusion. Whether the stimuli are measurable on ratio scales or only on nominal scales, the judgmental scale based on units of variability is approximately proportional to the logarithm of the scale constructed by one or another of the direct scaling methods. The extensive invariance of this logarithmic relation attests to a principle known throughout all of science—namely, that error or variability tends to be *relative*: the size of the error grows with magnitude. The principle finds expression under many phrasings: the standard deviation increases with the mean; the coefficient of variation remains constant; the signal-to-noise ratio stays put; accuracies are statable as one part in so many. The emergence of a similar canon in the subjective domain, a rule that variability tends to increase in proportion to the apparent magnitude, suggests an essential unity among the principles that govern quantitative relations in widely diverse endeavors.

For those who must build their science on one or another consensus of human judgment, a way seems open for an effective quantification.

References and Notes

1. L. L. Thurstone, "Psychophysical analysis," *Am. J. Psychol.* **38**, 368–389 (1927); *The Measurement of Values* (Univ. of Chicago Press, Chicago, Ill., 1959).
2. S. S. Stevens and E. H. Galanter, "Ratio

scales and category scales for a dozen perceptual continua," *J. Exptl. Psychol.* **54**, 377–411 (1957).
3. S. S. Stevens, "On the psychophysical law," *Psychol. Rev.* **64**, 153–181 (1957).
4. ———, "Problems and methods of psychophysics," *Psychol. Bull.* **55**, 177–196 (1958).
5. ———, "The surprising simplicity of sensory metrics," *Am. Psychologist* **17**, 29–39 (1962).
6. ——— and J. Volkmann, "The relation of pitch to frequency: A revised scale," *Am. J. Psychol.* **53**, 329–353 (1940).
7. F. M. Urban, *The Application of Statistical Methods to the Problems of Psychophysics* (Psychological Clinic Press, Philadelphia, 1908).
8. S. S. Stevens, "On the uses of poikilitic functions," in *Stimulus Generalization*, D. J. Mostofsky, Ed. (Stanford Univ. Press, Stanford, Calif., 1965), pp. 24–29.
9. E. Galanter and S. Messick, "The relation between category and magnitude scales of loudness," *Psychol. Rev.* **38**, 363–372 (1961).
10. L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (Univ. of Chicago Press, Chicago, Ill., 1929).
11. B. Finnie and R. D. Luce, "Magnitude-estimation, Pair-comparison and Successive-interval Scales of Attitude Items," *Dept. Psychol. Univ. Penn. Mem. MP-9* (1960). Finnie pursued the problem much further in a Ph.D. thesis entitled "An empirical comparison of magnitude and category scaling procedures applied to nonphysical stimuli," Harvard University, 1965. He compared magnitude and confusion scales for several kinds of nonmetric stimuli: attitude statements about the church, adjectives, adverbs, and occupational prestige, and statements expressing various degrees of cynicism. He also scaled the inverse continuum—for example, the degree of unfavorableness expressed by a statement. This work provides impressive new evidence to support the generality of the relation depicted in Fig. 3.
12. G. Ekman, "Discriminal sensitivity on the subjective continuum," *Acta Psychol.* **12**, 233–243 (1956); "Weber's law and related functions," *J. Psychol.* **47**, 343–352 (1959).
13. R. S. Harper and S. S. Stevens, "A psychological scale of weight and a formula for its derivation," *Am. J. Psychol.* **61**, 343–351 (1948).
14. G. Ekman and T. Künnapas, "Subjective Dispersion and the Weber Fraction," *Psychol. Lab. Univ. Stockholm Rept. No. 41* (1957); M. Björkman, "Some Relationships between Psychophysical Parameters," *Psychol. Lab. Univ. Stockholm Rept. No. 65*, (1958); "Variability Data and Direct Quantitative Judgment for Scaling Subjective Magnitude," *Psychol. Lab. Univ. Stockholm Rept. No. 78* (1960); H. Eisler, "Empirical test of a model relating magnitude and category scales," *Scand. J. Psychol.* **3**, 88–96 (1962).
15. S. S. Stevens, "Review of L. L. Thurstone: The measurement of values," *Contemp. Psychol.* **4**, 388–389 (1959).
16. T. Indow ["An example of motivation research applied to product design"], published in Japanese in *Chosa to gijutsu No. 102* (1961), pp. 45–60.
17. G. Ekman and T. Künnapas, "Note on direct and indirect scaling methods," *Psychol. Rept.* **6**, 174 (1960); "Measurement of aesthetic value of 'direct' and 'indirect' methods," *Scand. J. Psychol.* **3**, 33–39 (1962).
18. G. Ekman, "Two generalized ratio scaling methods," *J. Psychol.* **45**, 287–295 (1958).
19. ——— and T. Künnapas, "Scales of aesthetic value," *Percept. Motor Skills* **14**, 19–26 (1962).
20. S. D. Koh, "Scaling musical preferences," *J. Exptl. Psychol.* **70**, 79–82 (1965).
21. G. Ekman and T. Künnapas, "A further study of direct and indirect scaling methods," *Scand. J. Psychol.* **4**, 77–80 (1963).
22. S. I. Perloe, "The relation between category-rating and magnitude-estimation judgments of occupational prestige," *Am. J. Psychol.* **76**, 395–403 (1963).
23. S. S. Stevens and M. Guirao, "Subjective scaling of length and area and the matching of length to loudness and brightness," *J. Exptl. Psychol.* **66**, 177–186 (1963).
24. S. S. Stevens, "The measurement of loudness," *J. Acoust. Soc. Am.* **27**, 815–829 (1955).
25. T. Künnapas and I. Wikström, "Measurement of Occupational Preferences. A Comparison of Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 156* (1963).
26. T. Engen and D. H. McBurney, "Magnitude and category scales of the pleasantness of odors," *J. Exptl. Psychol.* **68**, 435–440 (1964).
27. G. Ekman and T. Künnapas, "Scales of conservatism," *Percept. Motor Skills* **16**, 329–334 (1963).
28. T. Künnapas and M. Sillén, "Measurement of 'Political' Preferences. A Comparison of Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 172* (1964).
29. S. S. Stevens, M. Guirao, A. W. Slawson, "Loudness, a product of volume times density," *J. Exptl. Psychol.* **69**, 503–510 (1965).
30. G. Ekman and T. Künnapas, "Scales of Masculinity and Femininity. A Further Study of Direct and Indirect Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 162* (1963).
31. G. Ekman, "Measurement of moral judgment," *Percept. Motor Skills* **15**, 3–9 (1962).
32. T. Sellin and M. E. Wolfgang, *The Measurement of Delinquency* (Wiley, New York, 1964).
33. D. Bernoulli, "Exposition of a new theory on the measurement of risk" (originally published in Latin in 1738), *Econometrica* **22**, 23–35 (1954).
34. S. S. Stevens, "Measurement, psychophysics and utility," in *Measurement: Definitions and Theories*, C. W. Churchman and P. Ratoosh, Eds. (Wiley, New York, 1959), pp. 18–64; E. Galanter, "The direct measurement of utility and subjective probability," *Am. J. Psychol.* **75**, 208–220 (1962).
35. S. S. Stevens, "A scale for the measurement of a psychological magnitude, loudness," *Psychol. Rev.* **43**, 405–416 (1936).
36. ———, "On the brightness of lights and the loudness of sounds," *Science* **118**, 576 (1953) (abstract).
37. P. L. Pradhan and P. J. Hoffman, "Effect of spacing and range of stimuli on magnitude estimation judgments," *J. Exptl. Psychol.* **66**, 533–541 (1963).
38. G. Ekman and L. Sjöberg, "Scaling," *Ann. Rev. Psychol.* **16**, 451–474 (1965).
39. The power functions in question were later published in S. S. Stevens, "The psychophysics of sensory function," in *Sensory Communication*, W. A. Rosenblith, Ed. (M.I.T. Press, Cambridge, Mass., 1961), pp. 1–33.
40. J. C. Stevens and M. Guirao, "Individual loudness functions," *J. Acoust. Soc. Am.* **36**, 2210–2213 (1964).
41. R. L. Hamblin, D. A. Bridger, R. C. Day, W. L. Yancey. "The interference-aggression law?," *Sociometry* **26**, 190–216 (1963).
42. This article (*Laboratory of Psychophysics Rept. PPR-306*) was prepared with support from NSF grant GB-3211 and NIH grant NB-02974. A slightly modified version will appear in D. K. Whitla, Ed., *Handbook of Measurement in Education, Psychology and Sociology* (Addison-Wesley, Reading, Mass., in press).

Science and the Problems of Development

H. J. Bhabha

This talk on science and the problems of development is based mainly on our experience in India during the last 20 years, but much of what we have to face here may be common to many underdeveloped countries, and some of them, I hope, will be able to profit from our experience, both successful and unsuccessful. It is interesting to note that practically all the ancient civilizations of the world—Persia, Egypt, India, and China—were in countries which are today underdeveloped. This is not surprising, since most of the countries of the world are still in this category. Western Europe is in fact a small area of the globe which outstripped the rest, essentially from the time of the Industrial Revolution, because of its development of modern science and the enhanced ability this gave it to utilize the forces of nature and to thus achieve a much higher material standard of life for its people. Western Europe was followed, and indeed to some extent overtaken, in industrial development by the United States, and more recently the category of the industrially developed countries has been joined by the Soviet Union, Japan, and a few others. A major part of the world, however, still remains underdeveloped by these standards. What the developed countries have and the underdeveloped lack is modern science and an economy based on modern technology. The problem of developing

Between 6 and 11 January, the Tata Institute in Bombay served as host to the International Council of Scientific Unions on the occasion of the latest of its biennial assemblies. Homi J. Bhabha, director of the Tata Institute and chairman of the Indian Atomic Energy Commission, addressed the assembly on the role which science can and should play in a developing country such as India. To those of us who knew Homi Bhabha and his dynamic nature, it was apparent that he was undertaking the task with even more than his normal vigor. During the address he spoke with a species of intensity that was unusual even for him.

At a reception later in the evening I asked if he would give me a copy of his manuscript for publication in *Science*. He expressed gratitude for the suggestion. His sudden death at the top of Mont Blanc two weeks later prevented him from reading the galley proof, but I am certain he would have wanted the address to appear in its current form.—FREDERICK SEITZ