VOL. 55, NO. 4 JULY, 1958

Psychological Bulletin

PROBLEMS AND METHODS OF PSYCHOPHYSICS[1]

S. S. STEVENS
Harvard University

The methods and procedures of psychophysics have been reviewed from time to time and marshaled into more-or-less logical array. If another such inventory is now in order it is because recent developments allow us to put the various methods in new perspective and to see more clearly how they articulate with the problems of psychophysics.

Let us admit first off that a concern with method is justified only if it leads to something beyond itself. The study of method, which I suppose is the proper meaning of that overworked term *methodology*, is one of those "necessary evils" whose justification lies in its potential contribution to the solving of substantive problems. But the problems are the main concern. If an empirical problem is worth solving, a method for it is worth developing, but it may turn out that there is little profit in fashioning tools to do what nobody wants done. Methodology can easily become methodolatry.

Psychophysical methods have at times been treated as though they were ends in themselves, and in many texts the term psychophysics has seemed to be synonymous with three "classical" procedures for solving an issue that few people care about. Little wonder then that psychophysics has sometimes been accused of inconsequence. Attitudes of this sort are not improved by the decision of a distinguished committee to define psychophysics as the use of a human observer as a "null instrument" to determine "equality or difference of sensations" (**20**, p. 59). According to the view of this committee, psychophysics is a strange land lying between "the physical and psychical" (**20**, p. 65). As Evans (**9**, p. 5) puts it, "Psychophysics at present, therefore, is limited to the *relative* evaluation of light beams with respect to normal observers under standardized conditions. . . ."

Psychophysics is really a much more nutritious subject than these conceptions imply. Seeking the laws that relate the responses of men and animals to the energetic configurations of the environment, it probes matters of deep human interest, and matters that often make a practical difference in the market place. For some of us, at any rate, a certain excitement attaches to the discovery that on "quantitative" or prothetic perceptual continua, such as brightness, loudness, heaviness, length, duration, etc., equal stimulus ratios produce equal sensation ratios (**31**). This principle means that the psychological magnitude is a power function

---

[1] Supported by a grant from the National Science Foundation and by Contract Nonr-1866(15) with the Office of Naval Research (Project Nr142-201, Report PNR-190). Reproduction for any purpose of the U. S. Government is permitted.

of the physical magnitude. And an example of practical utility can be seen in the application of this power law to the problem of predicting the loudness of a complex noise from physical measurements made on the spectrum of the sound (**29**, **33**).

Psychophysics has its problems and its methods. The purpose of this paper is to try to classify the methods in terms of the problems. It is a follow-up on some earlier attempts (**25**, **26**), in which a similar point of view was tried out in part, but in which the coverage was less systematic. A related effort, but rather different in outcome, was made by Guilford (**14**). In attempting an exercise of this sort we must realize that an element of arbitrariness attaches to taxonomies, and alternative schemes are always possible. Furthermore, we must forego the ambition to be exhaustive and completely consistent, for the methods of psychophysics are in a state of flux, and, as knowledge continues to expand, our understanding of procedures will change and improve.

The names that have become attached to the various methods show interesting vagaries, for in labeling our methods we fasten attention sometimes on one feature and sometimes on another. The name refers sometimes to the manner of presenting stimuli, sometimes to the task assigned the observer, and sometimes to the statistical treatment used to process the data.

Having names for methods provides a convenient shorthand for the description of experiments—as well as some handy items to ask about on examinations. But labeling is not without its drawbacks. Labeling produces jargon, and jargon leads to esoteric discourse. Many readers would find clarity improved if special names for procedures were banned and authors were forced to frame their descriptions in conventional English.

But names for the methods are probably here to stay, and our purpose will be to classify rather than to abolish. Actually, my greater interest is in the development of a schema by which the methods may be classified, rather than in any particular inventory of procedures, but the schema proposed can be illustrated by tables of methods. And since the methods exist to solve the problems of psychophysics, it is appropriate to comment on what certain methods may and may not achieve in the way of solutions.

THE PSYCHOPHYSICAL PARAMETERS

Psychophysics concerns the functional relation between stimulus and response: $R=f(S)$. This function is affected by numerous parameters. For the purpose of classifying methods it is convenient to distinguish three classes of these parameters, namely, the task undertaken by the observer, the manner in which the stimuli are presented, and the statistical measure employed in the description of the data. These classes and their principal subdivisions are listed in Table 1.

In Table 1 the various subdivisions of the three psychophysical parameters—task, stimuli, and statistic—are designated by a capital letter. These letters will be used in Table 2 to characterize the several psychophysical methods. But first let us examine the psychophysical parameters in a little more detail.

Task

The observer's task is normally set by means of instructions. The observer is "tuned" to react in one way rather than another, but from our present point of view only certain as-

TABLE 1

PSYCHOPHYSICAL PARAMETERS

Task of observer is to judge		Stimulus arrangement		Statistical measure	
C	Classification	F	Fixed	L	Measure of location (central tendency)
O	Order	A	Adjustable	V	Measure of variability or confusion
I	Intervals				
R	Ratios				
M	Magnitudes				

pects of this tuning are of consequence. It is important, for example, whether the observer is told to judge brightness, or hue, or saturation, but we are not here concerned with this aspect of the *Aufgabe*. Rather we take the attentional focus for granted and then ask what type of *relational* judgments the observer is trying to make. These relational judgments fall into five groups, as follows.

Classification (C). Here the observer's task is classification of one sort or another. He judges whether his perception meets some *nominal* criterion, with no reference to order among his perceptions. In the simplest case the observer, attending to some attribute or aspect of perception, judges whether it is *present* or *absent*. Thus he may press a key if he hears a tone, or if he hears a change in the tone. Or he may be required to say in what quadrant of a circle a light appeared, or in what interval of time a click was sounded. The task is that of detection, and the observer is set to behave as a yes-no device. In other cases the observer's task may be to judge *equivalence*, i.e., whether or not some criterion is met. The criterion may be set by a first stimulus, and the task may be to judge whether a second stimulus produces an equivalent effect, e.g., the loudness of one tone may be adjusted to match that of another tone. Or the criterion may be established by instruction, as when the subject is told to adjust a line to make it appear vertical, or to adjust wavelength to make a light appear pure green. Sometimes the classification problem is simply *identification* or *recognition:* is the present stimulus the same as some previous stimulus to which a name or number may have been assigned? Or, for example, was the sound presented an English word, and which word was it?

Order (O). The observer is set to judge greater or less, heavier or lighter, louder or softer, etc.

Intervals or distances (I). The observer's task is to judge apparent distance or difference between two or more perceptions. Ordinarily this takes the form of partitioning a continuum into apparently equal intervals or assigning stimuli to categories that seem equally spaced along a continuum.

Ratios (R). The observer attends to the relative magnitudes of two or more perceptions and reports the apparent ratios among them. Alternatively he may be set to produce stimuli that appear to stand in a prescribed ratio, which may be stated numerically or may be set in terms of some other pair of stimuli.

Magnitudes (M). The observer judges the apparent magnitude of a perception. He usually attempts to assign numbers proportional to the apparent magnitudes of a series of stimuli. Or he may be set to produce stimuli that correspond to a series of prescribed magnitudes.

Stimulus Arrangement

Since there is no end to the variety of procedures that may be used for the presentation of stimuli, it might appear that no useful criterion for classifying them is possible. On the other hand, there is an important procedural distinction between confronting the observer with a variable stimulus and confronting him with a fixed stimulus. Variable or adjustable stimuli are better for some purposes, fixed stimuli for others. As a practical matter, fixed stimuli can nearly always be used, but adjustable stimuli, unfortunately, are not always easy to come by. It is hard, for example, to devise continuously adjustable weights for a lifted weight experiment, or continuously adjustable concentrations for experiments on taste.

We will distinguish then between two stimulus arrangements:

Fixed stimuli (F). Fixed stimuli are those that are not varied during the time they are being observed. Usually, of course, they are varied between observations.

Adjustable stimuli (A). Adjustable stimuli are those that may be altered during the course of observation. Usually the observer does the adjusting by operating a control, but the experimenter may operate the controls, or the adjustments may be made automatically, as in the method we shall call "tracking."

Statistical Measures

How the data from a psychophysical experiment are processed usually depends on the experimenter's purpose. Neglecting the secondary frills of statistical descriptions we can divide the usual treatments into two classes, depending on whether the final measure used is one or another measure of location (or central tendency, so-called) or one or another measure of variability, confusion, or dispersion.

Measures of location (L). For most purposes the measure we want to use is either a mean or a median. We want to know the typical response of a subject or of a group of subjects. Given a measure of location, a measure of dispersion can then be used to gauge the precision of the judgments.

The choice of a proper measure of location often presents interesting problems whose solutions are far from obvious. For one thing, we have a choice among such conventional measures as the mode, median, arithmetic mean, geometric mean, and harmonic mean. But it may turn out that none of these is appropriate. The arithmetic mean is inappropriate, for example, when applied to the readings obtained with a particular instrument whose indications are a nonlinear function of the quantity we want to average. Elsewhere (28) the writer has tried to suggest how an iterative procedure might aid in the solution of some of these problems. For a variety of reasons, a defensible rule concerning measures of location in psychophysical experiments is simply: when in doubt use the median. The median has the advantage that it is invariant under nonlinear transformations so long as they are increasing and monotonic.

Measures of variability (V). For our present purpose the measures of variability include the conventional measures of dispersion (standard deviation, average deviation, interquartile range, etc.), as well as measures of confusion, such as the proportions of times a given stimulus is judged in different categories. Measures of variability are often used in the assessment of differential sensitivity or resolving power. They are also used, under various assumptions, as "distance" measures on psychological continua. This "unitizing" of dispersion may be legitimate on some types of continua, but on quantitative and intensitive continua the assumptions commonly made about discriminal dispersion are frequently in error (**31**). We will return to this problem later.

PROBLEMS AND METHODS

Some of the principal problems of psychophysics are listed in Table 2 along with some of the typical methods used in their solution. As is clear in Table 2, each of the problems of psychophysics may be regarded as one or another problem of scale construction construed in the widest sense of the term. This is scarcely surprising, for psychophysics, like most other parts of science, is mainly concerned with measurement. And since measurement is possible at different levels, ranging from nominal to ratio, the basic problems of psychophysics can be classified in a way that reflects these different levels.

In listing the problems in this way we must remind ourselves that they represent types or classes of issues that might concern the psychophysicist. Often the solution of one or another of these problems is not an end in itself, but is only a necessary step in the answering of a more far-reaching query. Especially in the more practical pursuits, such as human engineering, does it often turn out that the solving of problems like those in Table 2 is merely a means to an end. It has already been mentioned, for example, how the development of the sone scale of loudness has helped to answer a persistent problem facing the acoustical engineers (**29**, **33**). Even more extensive commercial applications are made of the Munsell color scales which are based on extensive psychophysical studies. Examples of this sort could be multiplied at length, but our interest here is more in fundamentals than in applications.

The methods listed under each problem do not exhaust the possibilities, nor do they all qualify necessarily as good procedures. On the other hand, these methods illustrate the procedures commonly used. The names given to the methods are mostly those in general use, although an occasional name is new, as is also an occasional method. In constructing tables of this sort one can scarcely avoid thinking of new procedures that ought to be put to test.

Let us now consider each problem in turn.

I. NOMINAL SCALES

The nominal scale is the most general type of scale. It is the primitive variety that involves only classification, with no ordering or metricizing. Perhaps under some definitions this simple process may qualify, not as measurement, but as a kind of halfway house on the road to it. On the other hand, there is no doubt that the psychological processes involved in the forming of classes, concepts, or categories present rich and varied problems. As a matter of fact, Bruner, Goodnow, and Austin (**7**) have written a whole book on the problem of "categorizing and conceptualizing," which is essentially a problem in nominal scaling. So too are the manifold problems in such areas as pattern recognition and articulation testing.

In what follows, we will limit our interest to the three instances of nominal scaling that have long been central problems in psychophysics and to a fourth problem whose develop-

TABLE 2

PSYCHOPHYSICAL PROBLEMS AND METHODS

I. To determine nominal scales	
a. Absolute thresholds	
1. Single stimuli	CFL
2. Counting	CFL
3. Forced location (forced choice)	CFL
4. Adjustment	CAL
5. Limits	CAL
6. Tracking	CAL
7. Staircase (up-and-down)	CFL
b. Resolving power or differential sensitivity	
1. Adjustment (average error)	CAV
2. Tracking	CAV-OAV-CAL
3. Constant stimuli	OFV
4. Single stimuli	IFV-OFV
5. ABX	CFV
6. Forced location	CFL
7. Quantal increments	CFL
c. Equation of magnitudes	
1. Adjustment	CAL
2. Constant stimuli	CFL-OFL
3. Tracking	CAL-OAL
4. Staircase (up-and-down)	CFL-OFL
d. Identification	
1. Single stimuli	CFV
II. To determine ordinal scales	
1. Pair comparison	OFL
2. Rank order (order of merit)	OFL
3. Rating scale	OFL-IFL
4. Single stimuli	CFV
III. To determine interval scales	
1. Equisection (bisection)	IAL-IFL
2. Interval estimation	IFL
3. Category rating (equal intervals)	IFL
4. Category production	IAL
5. Pair comparison	OFV
6. Rank order	OFV
7. Successive categories	IFV
8. Successive intervals	IFV
IV. To determine logarithmic interval scales	
1. Pair comparison	OFV
2. Ratio matching	RAL-RFL
V. To determine ratio scales	
1. Ratio estimation	RFL
2. Ratio production (fractionation, multiplication)	RAL-RFL
3. Magnitude estimation	MFL
4. Magnitude production	MAL

NOTE.—The capital letters after each method refer to the psychophysical parameters in Table 1. Alternative procedures under a given method are indicated by multiple sets of letters.

ment stems from information theory. Although conventional treatments do not usually subsume these problems under the heading of nominal scales, my colleague, Ulric Neisser, has pointed out to me that that is where they belong. Each of them reduces in one way or another to the classification of stimuli. Thus, in measuring absolute thresholds we form a twofold classification: those stimuli that can be perceived and those that cannot. Similarly, in measuring resolving power or differential sensitivity we divide stimulus increments into two classes, detectable and not detectable. In the equation of magnitudes our task is obviously to form the class of equivalent stimuli (according to a particular criterion) and to set if off from the class of stimuli that are not equivalent. And in the fourth problem our interest may be to determine into how many distinct classes a person can divide a set of stimuli without confusing any of them. Or it may concern the recognizability of stimulus configurations and the parameters that govern such recognition. In trying to solve these problems, we confront the observer with the task of putting stimuli into classes. The psychophysical procedure involved requires only classification and does not require that the observer order his perceptions or judge the intervals or ratios among them.

It may be true, of course, that under a particular experimental procedure the observer may be asked to judge "greater or less," but if the problem is really one of nominal scaling (e.g., the measurement of resolving power) the experimenter will proceed to use the experimental results to determine class boundaries. In other words, the categories of the nominal scale will be abstracted from the observer's judgments of apparent order.

Absolute Thresholds

The absolute threshold, the point that divides the continuum of stimuli into those the subject can detect and those he cannot, tends to elude our efforts to define its locus because it shifts about in time and we are forced to trap it by sampling and statistics. Mostly we use one or another version of the method of *single stimuli*. Fixed stimuli are presented at various levels and the yes-no responses of the subject are recorded. The class boundary which we call the threshold may be defined as the stimulus level detected half the time.

The method herein called *counting* is a procedure sometimes used in the large-scale testing of inexperienced subjects. Several stimulus levels are presented in a series and the subject reports how many he perceived. Other variations on this method may call for the presentation of two or more stimuli at a fixed level.

The method sometimes called *forced choice* (**5**, **40**) is not unlike counting except that, instead of asking the subject to say how many stimuli were presented, we ask him to say where in space or in which of several intervals of time the stimulus occurred. Since many methods call for a forced choice (e.g., constant stimuli used with two-category response), a better name for this procedure might be *forced location*. The subject tries to locate the stimulus in space or time. Blackwell has studied many of the parameters of this process.

The foregoing methods all make use of fixed stimuli. An adjustable stimulus can be used under the method of *adjustment*, in which the observer sets the level to be "barely detectable." This procedure is quick and convenient, but ordinarily it allows only a rough determination of the threshold.

If the level is varied systematically from points below and above threshold, and the subject signals when the threshold is crossed, we call the procedure the method of *limits*.

The term *tracking* is suggested to designate a procedure made popular by the Békésy audiometer (**2**, **3**). The observer presses a key whenever he hears a tone. As long as the key is pressed the level of the tone decreases, and when the key is released the level rises. By this means the subject may track his threshold throughout the frequency range, and by means of a recording pen the track of the stimulus is traced across an appropriate grid. A smoothed curve through the zigzag track is usually drawn to depict the threshold locus.

A method of tracking was apparently developed independently by Oldfield who used it for the measurement of visual thresholds (**18**). The method of tracking has also been used with animals. Blough (**6**) devised an experiment in which a pigeon was trained to peck at one key when a target was visible and at another key when the target was too dim to be seen. The pecks were made to control the brightness of the target, and the pigeon was thereby able to track its own dark-adaptation curve.

The *staircase* method, sometimes called the "up-and-down" method (**8**), is much like the method of tracking except that the stimulus level is not varied continuously. The levels used are fixed and discrete, and the level presented on a given trial depends on the response made on the previous trial. Thus if the previous stimulus was detected, the level of the next stimulus is lowered a step; if the previous stimulus was not detected, the level is raised.

Resolving Power

The measurement of the difference limen, ΔS, has been a major concern of classical psychophysics, and enormous labors have gone into the refinement of methods. Yet there is no agreed-upon best procedure for determining the least resolvable difference between two stimuli. Differential sensitivity is a difficult, "noisy" thing to measure, and it is not surprising that different procedures give different results.

Measures related to resolving power are sometimes obtained as a kind of by-product from such methods as *adjustment* and *tracking*. Some measure of dispersion about the mean adjustment or the "center" of the track may be used to define the just noticeable difference. In principle, measures of resolving power may be derived from the scatter of observations obtained from a wide variety of psychophysical procedures.

The method of *tracking* has also been adapted to the direct measurement of just noticeable differences (**19**, **42**). The trick here is to modulate the level of a signal alternately up and down, and to let the observer's responses control the amount of the modulation. He presses a key whenever he detects modulation and releases it when the level of the stimulus no longer appears to rise and fall. When the key is pressed the degree of modulation declines and when the key is released the degree of modulation increases. Thus the observer is able to "track" the just noticeable *change* in the signal. The average detectable modulation is determined by a curve drawn through the center of the zigzag track. This procedure has been used successfully with auditory stimuli, and there appears to be no reason why it should not lend itself equally well to some other types of stimuli.

Although it is not listed in Table 2, the staircase adaptation of the method of tracking could also be used to determine the threshold of a modulation.

Note that in Table 2 the method of tracking is scored CAV-OAV-CAL. This is intended to suggest that the observer may be instructed to judge in different ways, or that different statistical measures may be used to define the differential threshold. Other double scorings in Table 2 stand for alternative procedures under a given method.

The three methods, *constant stimuli*, *single stimuli* and *ABX*, all have much in common. Constant stimuli employs a standard and a series of comparison stimuli, and the observer judges ordinarily in terms of greater or less. Single stimuli (sometimes called absolute judgment) dispenses with the standard and the subject judges in terms of two or more categories, such as light, medium, or heavy. In the ABX method the subject reports (forced choice) whether the third stimulus is more like the first or the second of a series. In one study this method appeared to yield larger difference limens than the method of constant stimuli with a two-category forced-choice judgment (**21**).

Under the method of *forced location* increments are added to a steady stimulus and the subject is required to say when in time, or where in space, the increment occurred (**5**). The measure of sensitivity may be taken as the increment correctly identified half the time, after correction for chance.

The method of *quantal increments*, sometimes called the *quantal procedure* (**38**), attempts to measure the size of the stimulus increment needed to produce an all-or-none jump in effective excitation. As developed by Stevens and Volkmann (**39**), the procedure calls for a steady stimulus to which brief increments are added periodically. The observer presses a key whenever he perceives an increment. Under optimal conditions we obtain a rectilinear psychometric function of predictable slope, from which the size of the "neural quantum" can be gauged in terms of stimulus units. Whenever the possibilities for stimulus control are such that this method can be used, the writer thinks it is the preferred procedure. Psychometric functions of the predicted form have been obtained for pitch and loudness, and the data Mueller (**17**) obtained for brightness fit the linear functions predicted by the quantal hypothesis as well as, if not better than, they fit the sigmoid curves Mueller drew through them. And the slopes of some 45 psychometric functions all cluster about the predicted slope.

On the other hand, some experimenters have apparently not succeeded in reducing the variability and "noise," whether in the experimental procedure or in the observers, to the point of obtaining clear "quantal functions." The task of proving the stepwise character of discrimination is not easy. It is in some ways like trying to prove that the charge on the electron is constant. Prior to Millikan's oil-drop experiment, several attempts seemed to show that the charge was not constant but was probably normally distributed. For the electron the physicist now knows pretty well how to set up a repeatable experiment that will demonstrate the all-or-none nature of the charge, but as yet, unfortunately, we cannot prescribe all the conditions that will guarantee a "quantal" psychometric function. One reason is that the observer is too important a part of the specification. The writer can name several friends and colleagues who have been able to hold the steady attention needed for this task, but he does not know how to specify the differences between them and the observers who have not done so well at it.

Nevertheless, whether or not good clear quantal functions are obtained in any given experiment, the procedure itself has much to recommend it for the purpose of mapping resolving power. The method of quantal increments goes directly at the problem of what increments can be detected, and it provides for internal checks on the "noisiness" of the results obtained.

Equation of Magnitudes

The determination of equivalence is the problem we try to solve when we map such things as equal loudness contours, luminosity functions, contours of constant hue, etc. We try to determine the class of stimuli that appear equal with respect to a particular attribute. It resembles what the economist does when he maps indifference curves for utility (**35**).

For the typical matching problem the method of *adjustment* is usually the speediest and most straightforward. It is not, however, without its constant errors (**29**). The method of *constant stimuli* is also widely used for this purpose, and it too has its constant errors, in particular, the so-called "time-error" (**31**).

The method of *tracking* can also be used to trace out an equivalence contour. Zwicker and Feldtkeller (**41**) presented two tones alternately. One was of fixed intensity and frequency (1000 cps), and the other was made to sweep slowly through the frequency range. The observer pressed a key whenever the variable tone sounded louder than the standard and released the key whenever the variable sounded fainter. While the key was pressed, a motor-driven attenuator decreased the level of the variable tone, and when the key was released the motor reversed its direction and the variable grew louder. The data appear as a zigzag line traced by a pen across an audiometric chart, and an "average" line drawn through the zigzag tracing shows directly the form of the equal loudness contour.

The staircase or "up-and-down" method is an adaptation of the method of tracking for use with fixed values of the comparison stimulus. Applied to the problem of equating magnitudes, the staircase method is a kind of cross between constant stimuli and tracking. It is like constant stimuli in that a standard and a set of fixed comparison stimuli are used, but it is like tracking in that the response of the subject determines the value of the subsequent comparison stimulus. For example, if the subject says "greater" the next comparison stimulus is decreased; if he says "less" the next comparison stimulus is increased.

The mapping of invariances by means of a matching procedure is the basis of much of what we sometimes call measurement in psychophysics. When no other scales are available we often "measure" the effect of one factor in perception by changing that factor and then finding what alteration of another parameter will appear to undo the change. Thus we measure the effect of intensity on pitch by changing the intensity a given amount and then altering the frequency to restore the pitch to its original state. Or the observer may adjust the intensity instead of the frequency (**24**). We then measure the effect of intensity in terms of the frequency change required to cancel the effect of the change in intensity. Examples of this sort could be multiplied at length.

Identification

In the course of its burgeoning development, "communication theory" has had numerous impacts on psychology. Despite the fact that the theory deals only with the nominal properties of ensembles, not with their ordinal or interval properties, measures of information have found a use in many types of inquiry. And since the theory provides a mathematical model that concerns association at the nominal level of scales, it is not surprising that it should be put to work in psychophysics.

A typical problem in this area concerns the ability of observers to identify or recognize a stimulus in an absolute sense, without confusing it with other stimuli. The problem is related, of course, to differential sensitivity, but the concern is more with correct naming than it is with resolving differences. The distinction is something like that between pitch discrimination on the one hand and so-called absolute pitch on the other. Absolute pitch concerns the ability of a listener to name or identify the note played to him.

For many purposes it is important to know how many different stimuli (e.g., frequencies, intensities, colors, pressures, etc.) a person can identify with minimal error and to determine what factors affect information transmission on the different sensory continua (**1**). It turns out that, on many of the common sensory continua, perfect transmission of information (with no confusions) is ordinarily not possible with more than about five different stimuli (**16**). This fact is basic to many practical problems in the coding of information.

Also related to such problems is the question concerning the best way to distribute stimuli along a continuum in order to maximize information transmission. Studies of this sort have led to what is sometimes called an "equal discriminability scale" (**10**). Since the subject's task in these experiments is to identify and not to make comparisons, perhaps a better name would be "equal identifiability scale." In any case the psychophysical method employed in these problems is *single stimuli*. The experimenter presents a stimulus and the subject tries to give the appropriate response.

Problems of this sort are of course not limited to single stimulus dimensions. The problem of pattern recognition, for example, nearly always involves multivariate stimuli, usually visual or auditory although they may be tactual (**12**), and our interest is in how the observer is able to classify or recognize complex stimulus configurations. As already noted, it is to these problems that information theory has contributed useful tools.

II. ORDINAL SCALES

For the setting of perceptions in a rank order with respect to some aspect or attribute we have three conventional methods: *pair comparison*, *rank order*, and *rating scale*. The use of these methods for the purpose of ordering is usually straightforward and devoid of special problems. In the sensory area the ordering of psychological magnitudes is seldom a serious problem, because subjective magnitude is usually a monotonic function of stimulus magnitude, and we take the ordering for granted. Occasionally, however, we find that the simple monotonic relation breaks down. For example, the apparent saturation of a light of a single wavelength grows with intensity up to a certain value, but when the intensity is increased further the color blanches out and saturation declines.

The problem of ordering is sometimes extended to the relative *spacing* among stimuli, and the order of the spacing is sometimes deduced from confusions among the stimuli. The reasonable assumption is made that, if stimulus B is confused with A more often than C is confused with A, then C is farther from A than B is from A. The method of *single stimuli* is one procedure that might be used in an experiment of this sort.

Although measurements of confusion may order stimuli relative to a single point, as in the foregoing example, there are situations in which relative distances cannot be ordered by measures of confusion. On prothetic continua, where jnd's do not constitute subjectively equal distances (**31**) it is not always true that distances can be ordered by this procedure. Thus, if A and B are confused more often than C and D, it does not always follow that the distance from A to B is less than from C to D. In particular, if A and B are two intense tones, and C and D are two faint tones (all of the same frequency), tones A and B may be farther apart in subjective magnitude (sones) and still be confused more often. For example, tones of 100 and 101 sones would be confused more often than tones of 1.0 and 1.5 sones.

III. INTERVAL SCALES

Many interesting problems arise when we take on the task of erecting a scale of equal intervals on a psychological continuum. The determination of equal sense-distances seems to have originated with Plateau, who asked eight artists to paint a gray whose shade appeared equidistant between a black and a white (see **31**). In this manner, Plateau invented the method of *bisection*, which became *equisection* when the intervals were subdivided further. But Plateau, like many after him (see, for example, **13**), did not perceive the basic difference between bisection and such "ratio methods" as fractionation. As we shall see, this distinction is very important for two reasons: (*a*) bisection can lead at best only to an interval scale, and (*b*) it turns out that human observers are so constituted that they are generally unable to bisect an interval on a quantitative or intensitive (prothetic) continuum without making a systematic error.

The fact that bisection leads only to an interval scale is obvious enough. That subjects cannot perform valid bisections on certain types of continua is not so obvious, however. Evidence for this statement is described elsewhere (**31**), but since the argument is relevant to our present concern, let us review it briefly.

Systematic studies of more than a dozen perceptual continua have shown that these continua divide themselves into two varieties. Class I comprises the quantitative, intensitive continua, the continua concerned with *how much*. Discrimination on certain of these continua, such as loudness, brightness, and heaviness, seems to involve an additive process at the physiological level. For this reason Class I has been called *prothetic*. Class II includes the qualitative and positional continua, the continua concerned with *what* or *where*. Discrimination on these continua seems to involve a substitutive process, and they are therefore called *metathetic*.

Now, on prothetic continua we find that the psychological magnitude, as determined by ratio scaling procedures, approximates a power function of the stimulus magnitude. The rule is that equal stimulus ratios correspond to equal sensation ratios. This rule, the writer has suggested, is a basic "psychophysical law." The exponents of the power functions range from about 0.3 for loudness and brightness to about 3.5 for the apparent intensity of electric current applied to the fingers (**34**, **36**). Since differential sensitivity on these continua tends to approximate Weber's law, or rather the modified form of this law, $\Delta S = k(S+c)$, it follows that the psychological magnitude represented by ΔS increases as the stimulus increases. Discrimination, in other words, is not constant over the continuum, when measured in subjective units. There is a basic asymmetry in sensitivity.

On the other hand, on metathetic continua, discrimination, measured in subjective units, tends to be uniform over the scale, and there is no systematic asymmetry as there is with prothetic continua. This difference between the two continua makes for rather different behavior on the part of the observer. On the "symmetrical" metathetic continua, the results of bisection tend to agree with the results obtained by direct magnitude estimation and related procedures. But on the "asymmetrical" prothetic continua the point of bisection tends systematically to be lower than the point predicted by direct magnitude estimation. (Bisection is also plagued by a curious and dramatic order effect, which has been called "hysteresis" [**31**].)

The phenomena that characterize equisection also show up in the method of *interval estimation*, which is a kind of inverse of equisection. In equisection the intervals are adjusted to meet some criterion (usually equality), whereas in interval estimation the experimenter sets a series of stimuli and the observer estimates their apparent spacing. A convenient procedure for reporting these estimates is to have the observer adjust the positions of a set of markers along a line. The apparent intervals between successive markers are made to appear proportional to the apparent intervals between the stimuli. This method has been used with loudness and with lifted weights (**31**). The markers were movable sliders on a steel bar set before the observer. This procedure is in some ways analogous to the use of a continuous rating scale on which the judge places a pencil mark on a line. Since apparent position (on a line) is a metathetic continuum on which discrimination is not asymmetrical, adjustments of visual position provide an unbiased method of assessing the apparent spacing of other stimuli—except for possible distortions due to end effects (**37**). Thus the essential features that characterize equisection on prothetic continua (hysteresis and the bias due to the asymmetry of sensitivity) are also revealed by the method of interval estimation. (For an interesting variation on the method of interval estimation, see **17***a*).

The discrepancy between the "interval" judgment and the judgment of magnitude is especially striking when we use the method of *category rating*, under which the observer assigns a finite set of numbers or adjectives to a set of stimuli and tries to space the categories equally. Plotted against the ratio scale of subjective magnitude, the category scale has turned out to be concave downward on nine prothetic continua recently examined (**36, 37**). This nonlinearity in the category scales shows up even when the "pure" form of the category scale is obtained by a process of experimental iteration.

On metathetic continua, such as pitch, position, inclination, and proportion, the category scale *may be* linearly related to the magnitude scale, provided the distortions due to stimulus spacing, landmarks, and differential familiarity have been neutralized. These factors are some of the second-order variables that can alter the form of the category scale.

Another method for obtaining a category scale is the method of *category production* (**37**). This is a kind of inverse of category rating. Instead of asking the observer to assign categories to the stimuli, the experimenter names the categories, in irregular order, and the observer adjusts the stimulus to produce his conception of each category. Examples of the extreme categories (e.g., No. 1 and 7) may be presented to the observer at the outset. In our few tests of this method we found that it seemed to give directly a close approximation to the "pure" category scale that would be obtained by experimental iteration.

What is essentially the method of category production has also been used to study how people make linear interpolations in a spatial interval (**22**). This is a metathetic continuum. The experimenter is usually concerned with the "objective" accuracy of the observer's settings, although he may also be concerned with the form of the observer's subjective scale.

In summary, then, the four methods, equisection, interval estimation, category rating, and category production, are all designed to produce an interval scale of "equal sense-distances." If properly used, they can achieve this end on metathetic continua, but on prothetic continua they fail to produce intervals that are equal, as measured by the ratio scales of the continua.

We turn now to the class of methods that seek to produce equal intervals via the "unitizing" of one or another measure of variability. *Pair comparison* is perhaps the best known example of this procedure, but the underlying philosophy is similar for the other three methods listed in Table 2 (see **14**). By making certain simple assumptions regarding the distribution of the observed variabilities or confusions, we try to deduce the form of the underlying continuum. It would appear that on metathetic continua, where sensitivity to differences is uniform (in subjective units), the distribution assumptions most commonly invoked may lead to an interval scale. But on prothetic continua, the assumptions of normal and uniform variability are demonstrably in error, and therefore the resulting scales are not scales of equal intervals. On prothetic continua the procedures ordinarily used to derive equal intervals from measures of variability or confusion miss the mark for the same reason that "Fechner's law" fails: the subjective size of the jnd is not constant over the continuum. Likewise "discriminal dispersion" is not uniform over a prothetic continuum. As a matter of fact, the psychophysical power law, coupled with the relativity of resolving power, leads us to predict that discriminal dispersion is not distributed normally (or even symmetrically) on a linear subjective measure of a prothetic continuum, although it may be normal on a logarithmic measure of the continuum.

We must conclude, I think, that those procedures that make use of an assumed canonical distribution of variability are less useful for scaling than methods that utilize directly a measure of location. Even so, in the determination of equal intervals on prothetic continua, these latter methods are themselves subject to invalidating biases. It appears, therefore, that the only proper method for determining equal intervals on a prothetic continuum is to construct a ratio scale (see Section V). This solution is possible because the ratio scale contains the interval scale.

IV. LOGARITHMIC INTERVAL SCALES

The possibility that discriminal dispersion may increase proportional to the psychological magnitude on a prothetic continuum suggests that an assumption to this effect might make it possible to scale the continuum into intervals that are equal in terms of logarithms. So far as the writer is aware, no use has been made of such scales, although he has elsewhere (**31**, **35**) described some of their properties, including their mathematical group structure. In proceeding in this fashion we would be assuming that the conventional procedures used to scale a continuum by the method of pair comparison give us, for prothetic continua, a scale on which the values are separated not by equal intervals, but by equal ratios, i.e. $a/b=b/c=c/d= \ldots$.

The equating of ratios, either by way of a processing of variability or by a direct judgment of apparent ratios, would provide the basis for what the writer has called a *logarithmic interval* scale. This scale is invariant under a power transformation, i.e., for any value x we can substitute x' where $x'=ax^b$, and where a and b are positive numbers. As with the linear scale of equal intervals, the zero point on the logarithmic scale can be chosen arbitrarily and moved at will.

If such a scale were desired, the straightforward procedure for achieving it would presumably be some procedure of direct *ratio matching*. Methods of this type have not been very thoroughly explored, although Garner (**11**) tried equating loudness ratios. He seemed to find evidence that observers may not be able to keep separate the two tasks, that of equalizing ratios and that of equalizing intervals. On the other hand, J. C. Stevens obtained interesting results when two brightnesses were used to define a subjective ratio, and the observer adjusted the ratio between two loudnesses to make the ratio between the loudnesses match the apparent ratio between the brightnesses (**31**).

Ratio matching of this kind may have utility for psychophysics, because it provides an alternative method for demonstrating that the "psychophysical law" governing prothetic continua is a power function, i.e., the psychological magnitude is equal to the stimulus magnitude raised to a power. In principle, the power function can be tested by this method without requiring the observer to make numerical estimates of ratios or of magnitudes (as described in the next section). The observer is required only to make the apparent ratio between one pair of stimuli equal the apparent ratio between another pair of stimuli. Then if it turns out that the ratio of one pair always equals the ratio of the other pair raised to a power, it follows that the psychological magnitudes are power functions of their respective stimuli.

For example, suppose the subject adjusts pairs of luminances (B_1 and B_2) to make the ratio of the brightnesses equal the ratio of the apparent lengths of two lines (L_1 and L_2). If for all possible pairs so matched we find

$$L_1/L_2=(B_1/B_2)^n$$

where n is a constant, then both subjective length and subjective brightness are power functions of their respective stimuli. The limitation on this approach is that we could not by this procedure determine the value of the two exponents involved. That would require the additional methods described below. We could, however, by such ratio matchings determine the *relative* values of the exponents for length and brightness.

Whether observers can make such ratio matches with sufficient consistency to make it a profitable procedure has not been explored very thoroughly. Nevertheless, the experiments involving the matching of ratios of loudness and brightness have thus far been encouraging. In this particular case, the observers adjusted the two sounds to approximately the same physical ratio as the experimenter set between the two luminances. This outcome is consistent with other evidence that loudness and brightness are both power functions of their stimuli, not logarithmic functions as Fechner supposed, and that the exponents for both loudness and brightness are approximately the same. By means of the ratio scaling procedures discussed below we have shown that both exponents are of the order of 0.3 (**31**).

V. Ratio Scales

Perhaps the lack of interest in logarithmic interval scales stems from the fact that the scientist's greater interest lies in ratio scales. He is less interested in ratios whose values are equal but indeterminate than he is in ratios whose values he can specify. If he has a procedure for equating ratios, plus a procedure for equating intervals, he can proceed to construct a ratio scale on which the zero point is not arbitrary (**26**).

We have seen, however, that observers exhibit a systematic bias when they try to equate intervals on a prothetic continuum. Whether, and under what conditions, observers can equate a series of unknown ratios has not been fully explored. In view of this state of affairs, how do we create ratio scales on a perceptual continuum?

The answer seems to be that we ask the subject to judge the value of the ratio, or of the magnitude, directly. Using one or another, or preferably combinations, of four different methods we proceed directly to the goal of assessing relative psychological magnitudes. The potential methods are as follows:

Ratio estimation calls for the presentation of two or more stimuli and the observer names the value of the apparent ratio between them. The so-called *constant sum* method (**15**) is a special instance of this procedure. As typically used, this method requires that the observer divide 100 points between two stimuli in such a way that the division between the points reflects the apparent ratio between the sensations. In general, however, there is little reason to restrict the observer's method of report in this manner. Restrictions and constraints on the observer are often a source of trouble and bias in ratio and magnitude judgments.

It should be mentioned that, in using any of the methods for determining ratio scales, the experimenter will generally do well to compute medians, for there is no limit to how far an occasional observer may deviate from the rest of the group.

Ratio production is probably best known by the name of one of its subvarieties, *fractionation*. The observer adjusts a stimulus to produce a prescribed ratio between two apparent magnitudes. He sets a variable to be $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, or some other fraction of a standard. Alternatively the experimenter may set the stimuli and the observer may report whether they meet the criterion of a prescribed ratio. This procedure is analogous to the method of constant stimuli and is sometimes called by this name. The principal drawback of fractionation with fixed stimuli is that the choice of the levels at which the comparison stimuli are set may be critical. This difficulty can presumably be surmounted if the spacing of the comparison stimuli is determined by a process of experimental iteration under which the levels are altered in accordance with the outcome of successive experiments in such a way that the "criterion" stimulus is made to lie at or near the center of the range of the comparison stimuli (**37**).

The method of *fractionation* has its inverse in the method of *multiplication* under which the observer sets a variable to some prescribed multiple of a standard. In order to balance out certain sources of bias it is nearly always wise to complement fractionation with multiplication (**32**). This is especially true when wide ranges of stimulus values are explored in experiments on loudness and brightness.

Magnitude estimation refers to a procedure by which the observer makes a direct numerical estimation of the psychological magnitudes of a series of perceptions (**30**). Two main varieties of this procedure have been used. Under one of them the experimenter presents a stimulus and assigns it a number (modulus) such as 10, say. He then presents other stimuli, and the observer assigns to them numbers proportional to their apparent magnitude. Under the other variation in the method, no modulus is prescribed. The stimuli appear in irregular order and the observer assigns numbers proportional to magnitude, using a modulus of his own choosing.

The method of direct magnitude estimation has given good results with loudness, brightness, lifted weights, duration, lightness of grays, visual length, pitch, proportion (**37**), finger span (unpublished), vibration (unpublished), and electric shock (**34**, **36**). Incidentally, with this method it often turns out that the geometric mean falls close to the median.

Magnitude production is a method that has been named but not yet thoroughly explored (**31**). It is the inverse of magnitude estimation in that the experimenter states a series of magnitudes (presumably in irregular order) and the observer adjusts a stimulus to produce them. In some ways the procedure resembles the method of category production described above, except that no range is specified and the observer tries to judge in terms of apparent magnitudes and not in terms of a finite number of prescribed categories. Magnitude production is a potentially interesting method, provided the stimulus control is such that the subject can adjust the stimulus over the required range, but many questions concerning its peculiarities and difficulties remain to be answered. It is not unlikely that the biases in magnitude production are such that, in a balanced program, they might be used to offset some of the systematic errors in magnitude estimation. It seems, in general, that each of the ratio-scaling methods may contain biases peculiar to itself, and that the elimination of the biases can sometimes be achieved by means of a counterbalanced design in which the biases inherent in one method are evaluated and corrected by means of a method that contains biases of an opposite sort. The principle is analogous to that employed in the use of a balance in weighing an object: in order to discover and correct for possible asymmetries in the balance we interchange the weights on the scale pans.

We have already noted that, on prothetic continua, ratio scales of psychological magnitude turn out to be power functions of the stimulus magnitude. The power law seems to hold on at least 16 perceptual continua. On metathetic continua, the ratio scaling methods may or may not give a power function. When a power function is found to hold, the exponent appears to be 1.0. But on a metathetic continuum like pitch, the psychological magnitude, measured in mels, is definitely not a power function of frequency in cycles per second (**37**).

Another important point to note is that, when prothetic continua are scaled by the ratio methods, Fechner's law obviously fails. The relation between perceptual magnitude and stimulus magnitude is not logarithmic. Even if we take the weaker form of Fechner's law, which says that the counting off of jnd's gives the scale of perceptual magnitude, we find that the law fails almost as badly. Successive jnd's, in other words, are not subjectively equal.

On metathetic continua, on the other hand, the jnd scale apparently coincides with the scale of psychological magnitude. Thus, to a fair approximation, jnd's in pitch represent constant increments in mels (**27**).

THE ESTIMATION OF OBJECTIVE VALUES

The methods of psychophysics are ordinarily designed to solve problems related to the nature of organisms. The focus of interest is typically the normal observer, his thresholds, his resolving powers, and the magnitudes of his perceptions. The methods have, of course, their clinical uses, and the assessment of individual differences is central to many practical undertakings.

Quite different in aim, but sometimes similar in procedure, is another human activity involving discrimination and judgment. This is the use of the human being as an instrument to measure the objective values of things. Despite the ingenuity of modern instrumentation, many tasks of rating, grading and judging can still best be done by two-legged meters (cf. **23**). Man's sensing, differentiating, and integrating circuits still surpass in flexibility and power any inanimate substitutes yet devised. Instruments may aid but they do not displace the wine taster, the leather grader, the lumber sorter, or any of a host of other judges on whom commerce depends for the appraisal of its wares. Little of this type of activity gets attention in the academic laboratory, although much could probably be learned from its systematic study.

In the framework of our present concerns, the assessing and grading of objective things are practical problems—a potential field perhaps for an applied psychophysics. The chief difference between problems of this type and those we have been considering lies in the point of view. In addressing any of the five problems in Table 2 we seek to learn the properties of the human instrument; in problems of grading we care nothing about the properties of the instrument as such, but only about the accuracy of its indications. In the grading of wool, for example, the mill owner hopes the assessment of the "clip" will tell him more about the wool than it does about the grader. He hopes in other words, that the grader will commit the "stimulus error" 100%. The psychophysicist presenting a series of tones to be judged for loudness hopes quite the opposite. He wants the subject to report apparent loudness and not to judge how many decibels are probably being produced by the earphone. Some experienced judges can do either task at will, but the estimation of decibel levels requires a lot of training. How the properties of the judge, as appraised and systematized by psychophysics, interact with the applied problems of grading and rating is a potentially interesting problem.

The capacity of the human instrument to make correct assessments of this or that is also a central issue in some branches of engineering psychology, and it is in these connections that the most systematic studies have been made. This is not the place to review this lively field, but an illustrative example may be in order.

It concerns the continuous control of a complex system, like the problem faced by the pilot taking a large ship through the narrow channels of the Suez Canal. The same kind of problem comes home to many of us when we try to back an automobile with a luggage trailer tied on behind. We watch the trailer to see where it is going. It goes to the left, let us say, when it should go to the right, so we turn the front wheels of the car and watch to see what happens. The trailer corrects its course to the desired direction, whereupon we straighten out the front wheels and continue to back up. To our dismay the trailer keeps turning to the right. We find we have overcorrected, because we attended only to the "output" of the system, and did not take into account the delay between the input control (turning the front wheels) and the final movement of the trailer. With enough practice, a driver can learn to manage the procedure and even to back the trailer into a garage. He must learn not only to judge the position and direction of the trailer, but also to act as an integrator and predict the effects of his control actions as they will be summed up over a period of time.

Problems of this sort have opened new fields of study concerned with man-machine control systems (**4**). One of the primary problems is to learn what objective aspects of the situation the operator can judge with greatest reliability, and then to display to him only those features that match his judgmental capacity. It has been found, for example, that best results are achieved when the human operator is relieved of the task of performing integrations and differentiations. The controls and displays must be engineered in such a way that the operator can effectively control the system by acting as a simple amplifier. The need for a complex and difficult judgment of objective values is thereby replaced by a simpler demand on judgment.

REFERENCES

1. Alluisi, E. A. Conditions affecting the amount of information in absolute judgments. *Psychol. Rev.*, 1957, **64**, 97–103.
2. von Békésy, G. Über ein neues Audiometer. *Arch. elekt. Übertragung*, 1947, **1**, 13.
3. von Békésy, G. A new audiometer. *Acta. Oto-Laryng. Stockh.*, 1947, **35**, 411–422.
4. Birmingham, H. P., & Taylor, F. V. A design philosophy for man-machine control systems. *Proc. I.R.E.*, 1954, **42**, 1748–1758.
5. Blackwell, H. R. Psychophysical thresholds. *Engng Res. Bull.* No. 36, Engineering Research Institute, Univer. of Michigan, 1953.
6. Blough, D. S. Method for tracing dark adaptation in the pigeon. *Science*, 1955, **121**, 703–704.
7. Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.
8. Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
9. Evans, R. M. *An introduction to color.* New York: Wiley, 1948.
10. Garner, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.
11. Garner, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1954, **26**, 73–88.
12. Geldard, F. A. Adventures in tactile literacy. *Amer. Psychologist*, 1957, **12**, 115–124.

13. Graham, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.

14. Guilford, J. P. *Psychometric methods* (2nd ed.) New York: McGraw-Hill, 1954.

15. Metfessel, M. F. A proposal for quantitative reporting of comparative judgments. *J. Psychol.*, 1947, **24**, 229–235.

16. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

17. Mueller, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. *J. gen. Physiol.*, 1951, **34**, 463–474.

17*a*. Newhall, S. M. A method of evaluating the spacing of visual scales. *Amer. J. Psychol.*, 1950, **63**, 221–228.

18. Oldfield, R. C. Continuous recording of sensory thresholds and other psychophysical variables. *Nature*, 1949, **164**, 581.

19. Oldfield, R. C. Apparent fluctuations of a sensory threshold. *Quart. J. exp. Psychol.*, 1955, **7**, 101–115.

20. Optical Society of America, Committee on Colorimetry. *The science of color.* New York: Crowell, 1953.

21. Rosenblith, W. A., & Stevens, K. N. On the DL for frequency. *J. acoust. Soc. Amer.*, 1953, **25**, 980–985.

22. Schubert, R. G., & Jenkins, W. L. The effect of brief training on linear interpolation. *J. appl. Psychol.*, 1956, **40**, 53–54.

23. Scott Blair, G. W. *Measurements of mind and matter.* New York: Philosophical Library, 1956.

24. Stevens, S. S. The relation of pitch to intensity. *J. acoust. Soc. Amer.*, 1935, **6**, 150–154.

25. Stevens, S. S. Sensation and psychological measurement. In E. G. Boring, H. S. Langfeld, and H. P. Weld (Eds.), *Foundations of psychology.* New York: Wiley, 1948.

26. Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

27. Stevens, S. S. Pitch discrimination, mels, and Kock's contention. *J. acoust. Soc. Amer.*, 1954, **26**, 1075–1077.

28. Stevens, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.

29. Stevens, S. S. The calculation of the loudness of complex noise. *J. acoust. Soc. Amer.*, 1956, **28**, 807–832.

30. Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

31. Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

32. Stevens, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.*, 1957, **29**, 603–606.

33. Stevens, S. S. Calculating loudness. *Noise Control*, 1957, **3**, No. 5, 11–22.

34. Stevens, S. S. Measurement and man. *Science*, 1958, **127**, 383–389.

35. Stevens, S. S. Measurement, psychophysics, and utility. In *Symposium on measurement: held by the AAAS, December 1956.* New York: Wiley, in press.

36. Stevens, S. S., Carton, A. S., & Shickman, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, in press.

37. Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, **54**, 377–411.

38. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–355.

39. Stevens, S. S., & Volkmann, J. The quantum of sensory discrimination. *Science*, 1940, **92**, 583–585.

40. Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

41. Zwicker, E., & Feldtkeller, R. Über die Lautstärke von gleichförmigen Geraüschen. *Acustica*, 1955, **5**, 303–316.

42. Zwicker, E., & Kaiser, W. Der Verlauf der Modulationsschwellen in der Hörfläche. *Acustica* 2, *Akust. Beih.*, 1952, **4**, AB 239–246.

Received January 2, 1958.