On scales of sensation:
Prolegomena to any future psychophysics
that will be able to come forth as science*

LAWRENCE E. MARKS†

John B. Pierce Foundation Laboratory and Yale University, New Haven, Connecticut 06519

Sensory scales fall into two classes. Type I scales of sensory intensity can be approximated by metric scaling procedures (magnitude estimation, magnitude production) and nonmetric procedures (conjoint measurement); Type I scales are supported by theoretical consideration of sensory processes. Type II scales of sensory dissimilarity can be approximated by metric scaling procedures (category rating, interval estimation, equisection) and nonmetric procedures (analysis of proximities). The psychophysical functions that relate Type I and Type II scales to their corresponding physical scales are in both cases power functions, but the exponents that govern Type I functions are typically about twice as large. Both Type I scales of sensory intensity and Type II scales of sensory dissimilarity are meaningful measures of perceptual experience, but they are measures of different aspects of perception. The duality of sensory scales helps to explain some apparent contradictions among divergent attempts to validate scales of sensation.

There is a certain respect in which I write this paper in sadness. It is regrettable that, as we approach the last quarter of the 20th century, a fundamental debate still rages about the meaningfulness, necessity, and validity of scales of sensation, a debate hardly different in its essence from that which centered around Fechner's work a century ago. For even though enormous talent, energy, and time has gone into the study of psychophysical scaling, nevertheless opinion is far from unanimous. The purpose here is neither to rehash old theories nor to review stale philosophical viewpoints. Rather, the aim is to present a reasonably simple, coherent, and consistent account of psychophysical scales of sensation that does justice not only to several varieties of scaling data, but also to certain facts about sensory functioning that themselves need not depend on scaling. What I hope to show is not only that there can be meaningful systems of subjective scales, but that the scales—or at least some of them—may even be necessary to any comprehensive understanding of sensory and perceptual processes.

The present paper begins by examining scaling procedures that require subjects to make direct, quantitative judgments of sensory magnitudes. Results obtained by these procedures pose a problem, in that data appear not to yield convergent or invariant

scales. Different classes of procedure—ratio and interval scaling—give results systematically at odds with each other; moreover, even within a given class of procedures (indeed, within a given method) results vary significantly from one study to another. The present theory postulates that there exist, for any given sensory attribute, two basic underlying scales: one a scale of magnitude, the other a scale of dissimilarity. Typically, these two scales are nonlinearly related. In order to establish the existence and nature of the underlying scales, we turn to consideration of sensory processes and to nonmetric properties of scaling data. Once some basic scales are determined, it becomes possible to return and reconsider data obtained through metric scaling procedures. *On the average*, results obtained by ratio-scaling procedures *approximate* the underlying scales of sensory magnitude, and those obtained by interval-scaling procedures *approximate* the underlying scales of dissimilarity.

SCALES DERIVED FROM DIRECT JUDGMENT

Ratio-Scaling Procedures

The point of departure for evaluating sensory scales is the work that involves direct judgments of sensory magnitudes. As a matter of convenience, I shall stick to the common terminological distinction between ratio-scaling and interval-scaling procedures. The most famous of the ratio-scaling procedures is Stevens's (1956) method of magnitude estimation, whereby subjects attempt to gauge the strength of their sensations by assigning numbers in proportion to

sensory magnitudes. Stevens and his colleagues derived scales of sensation for a large number of sensory and perceptual continua. On most if not all of these continua, the relation between sensation and stimulus turned out to approximate a power function (Stevens, 1957, 1961a). As long as parameters of stimulation other than intensity are held constant, the exponent of the power function is said to be a roughly constant expression of the rate of growth of sensation magnitude. Average values of reported exponents range from one-third (brightness of lights not too small in size or short in duration) to greater than three (electric shock to the fingers).

. Actually, the summary just given is an oversimplification. Even when stimulus conditions are held constant, there is a sizeable variation, from experiment to experiment, in the value of the power-function exponent obtained on any given modality. Exponents depend, for example, on: (1) stimulus range—usually, the smaller the range, the larger the exponent (Stevens, 1956; Jones & Woskow, 1966); (2) procedure—magnitude production, in which subjects adjust stimulus intensity to match numbers given by the experimenter, produces larger exponents than does magnitude estimation (Stevens & Greenbaum, 1966); (3) individual differences—when results are obtained on individual subjects, given otherwise identical experimental conditions, the exponent is not constant, but can vary over a range of at least two-to-one (Marks & J. C. Stevens, 1966). For an extensive review of the literature on these variations, see Poulton (1968). Table 1 and Fig. 1 display exponents of power functions obtained from ratio-scaling procedures applied to loudness. The range of exponents is about two-to-one.

Fig. 1. Exponents of psychophysical power functions for loudness. Circles: exponents derived from interval-scaling experiments. Squares: exponents derived from ratio-scaling experiments.

Interval-Scaling Procedures

There is the second class of experimental procedures, ones that produce another set of psychophysical functions, different from those produced by ratio-scaling procedures. In this second class of procedures, subjects are called upon to judge the sizes of sensory intervals or differences. The task may be to bisect an interval—to set one stimulus so as to make its sensation appear to lie halfway between two others; it may be to rate stimuli on a scale from 1 to 7 so that each number marks off a constant increment in sensation magnitude; it may be to estimate numerically the subjective size of the difference between two stimuli.

Typically, the results of experiments that employ interval-scaling procedures are also consistent with psychophysical power functions, but yield exponents smaller in size than those produced by ratio-scaling procedures (Stevens, 1961b; Marks, 1968). Furthermore, as with the results of ratio scaling, several procedural variables influence the size of the obtained power-function exponent (Marks, 1968). Table 1 and Fig. 1 display exponents of power functions obtained from interval-scaling procedures applied to loudness. Here the range is about three to one.

So we are faced with a double predicament. On the one hand, we find that no given type of procedure yields a wholly invariant set of psychophysical functions, and, therefore, no invariant set of sensory scales; furthermore, we find two somewhat different types of procedure that yield systematically different sets of functions. As Table 1 and Fig.1 show for the attribute of loudness, there are two overlapping distributions of power-function exponents, one derived from experiments that employ ratio-scaling procedures, the other interval-scaling procedures. The median value of the exponent is 0.54 in the former case, 0.27 in the latter, when the stimulus is measured in terms of sound pressure. Thus the ratio of the median exponents turns out to be exactly two to one.

Which Scales, If Any?

Herein lies the rub. It would appear that we must conclude either that the sensory system changes from procedure to procedure as well as from condition to condition, from experiment to experiment, from subject to subject, or alternatively, that sometimes, at least, in the course of making quantitative judgments, subjects do not apply numbers or numerical concepts in a linear fashion. Rarely before has the first alternative been invoked, nor will it now. Thus we are left with the second. Stevens (e.g., 1971) argued that a balanced average of a large number of experiments using ratio-scaling procedures can yield valid scales and valid psychophysical functions. Anderson (e.g.,
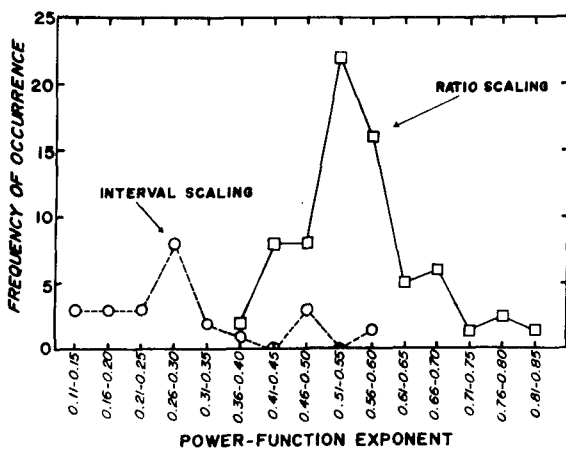
Table 1
Exponents of Psychophysical Power Functions for Loudness

Experimenter	Procedure	Parameters	Exponent
Ratio-Scaling Experiments			
Richardson & Ross, 1930	Magnitude Estimation	550-1100 Hz	0.44
Ham & Parkinson, 1932	Magnitude Estimation, Ratio Production	260-2500 Hz	0.67
Geiger & Firestone, 1933	Ratio Production	1000 Hz	0.85
Churcher, King, & Davies, 1934	Ratio Production	800 Hz	0.53
Rschevkin & Rabinovitch, 1936	Ratio Production	1000 Hz	0.44
Stevens, 1956	Magnitude Estimation	1000 Hz	0.62, 0.67, 0.55, 0.44, 0.55, 0.54, 0.80, 0.57, 0.44, 0.50, 0.60, 0.60
Stevens & Poulton, 1956	Fractionation	1000 Hz	0.65
Stevens, 1957	Ratio Production	1000 Hz	0.60
Robinson, 1957	Ratio Production	1000 Hz	0.50
J. C. Stevens & Tulving, 1957	Magnitude Estimation	1000 Hz	0.70, 0.56
J. C. Stevens, 1958	Magnitude Estimation	Noise	0.60
Scharf & J. C. Stevens, 1959	Magnitude Estimation, Magnitude Production	1000 Hz	0.60
Feltkeller, Zwicker, & Port, 1959	Ratio Production	1000 Hz	0.56
Reynolds & Stevens, 1960	Magnitude Estimation	Noise	0.56 Binaural 0.50 Monaural
	Magnitude Estimation	Noise	0.67 Binaural 0.63 Monaural
	Ratio Production	Noise	0.55 Binaural 0.58 Monaural
Stevens & Guirao, 1962	Magnitude Estimation	Noise	0.52
	Magnitude Production	Noise	0.70
	Magnitude Estimation	1000 Hz	0.43
	Magnitude Production	1000 Hz	0.67
Eisler, 1962	Magnitude Estimation	Noise	0.52
Schneider & Lane, 1963	Magnitude Production	Noise	0.70
Hellman & Zwislocki, 1964	Magnitude Estimation, Magnitude Production	1000 Hz	0.53 0.57
J. C. Stevens & Guirao, 1964	Combined Estimation-Production	1000 Hz	0.77, 0.69
J. C. Stevens & Hall, 1966	Magnitude Estimation	Noise, Durations 5-500 msec	0.54, 0.54, 0.57, 0.54, 0.53, 0.54, 0.52, 0.57, 0.52, 0.55
Hellman & Zwislocki, 1968	Magnitude Estimation, Magnitude Production	1000 Hz	0.54
Richards, 1968	Fractionation	1000 Hz	0.50
Rowley & Studebaker, 1969	Magnitude Estimation, Magnitude Production	1000 Hz	0.48 0.56
Scharf & Fishken, 1970	Magnitude Estimation	1000 Hz	0.43 Binaural 0.48 Monaural
	Magnitude Production	1000 Hz	0.53 Binaural 0.52 Monaural
Dawson, 1971	Magnitude Estimation	1000 Hz	0.53
Schneider, Wright, Edelheit, Hock, & Humphrey, 1972	Magnitude Estimation	200 Hz	0.55
		250	0.58
		300	0.58
		350	0.52
		400	0.50
		600	0.48
		800	0.37
		1100	0.42
		1600	0.40
		2500	0.42
Cross, 1973	Magnitude Estimation	Noise	0.64
Interval-Scaling Experiments			
Garner, 1954	Equisection	1000 Hz	0.26
Stevens, 1955	Bisection	1000 Hz	0.36 Average
H. Rubin (Data reported by Stevens & Galanter, 1957)	Category Rating	Noise	0.15, 0.18, 0.20, 0.25, 0.26, 0.29
Garner (Data reported by Stevens & Galanter, 1957)	Category Rating	1000 Hz	0.16
Stevens & Galanter, 1957	Category Production	Noise ?	0.28
	"Pure" Category Rating		0.26

Table 1 Continued

Experimenter	Procedure	Parameters	Exponent
J. C. Stevens, 1958	Category Rating	Noise	0.13, 0.33
Stevens & Guirao, 1962	Category Production	1000 Hz	0.30
	Equisection		0.50
Eisler, 1962	Category Rating	Noise	0.14
Schneider & Lane, 1963	Category Production	Noise	0.18
Beck & Shaw, 1967	Interval Estimation	1000 Hz	0.30, 0.31
Dawson, 1971	Interval Estimation	1000 Hz	0.58
	Dissimilarity Estimation		0.46
	Subtracted-Difference Estimation		0.46
Ward, 1972	Category Rating	1000 Hz	0.39
Parker & Schneider, 1974	Interval Estimation	1200 Hz	0.27
	Similarity Estimation	1200 Hz	0.24

1972) argued that under specific experimental conditions, rating procedures can yield valid scales and valid psychophysical functions. An even more skeptical view is that neither type of procedure yields valid scales. Instead, I suggest a more optimistic view— namely that both ratio-scaling procedures and interval-scaling procedures often approximate valid, though usually different, scales of sensation.

The double dilemma noted above requires that the present theory of sensory scales deal with two questions. First, how can we demonstrate that there exist two valid sets of scales of sensation? Second, how can we decide when a particular scaling procedure has yielded a valid sensory scale? The answers to these two questions are not totally independent. Independent formulation, however, is useful from a pedagogic point of view.

The present theory answers both questions by going beyond scaling experiments alone, by integrating data derived via scaling with sensory and perceptual data that do not rely on scaling procedures. In doing this, we must examine sensory scales from the points of view of internal consistency within a given sense modality, appropriate transitivity across modalities, and the enmeshing of fact and theory. When we do this, we come out with the two types of scale, which may be termed most neutrally as Type I and Type II. The set of Type I scales are what traditionally have been thought of as scales of sensory magnitude; these are the scales that often may be approximated through the use of ratio-scaling procedures. The set of Type II scales are perhaps best characterized as scales of sensory dissimilarity; they can often be approximated by the use of some of the interval-scaling procedures, among which I include judgments of sensory similarity and dissimilarity per se.

Before entering the mainstream of the arguments, it is worthwhile to make some terminological distinctions, which are of heuristic value, among three types of quantitative relation. Specifically, let us distinguish what I have elsewhere (Marks, 1974) called psychophysical, sensory-physical, and psychosensory relations.

Psychophysical relations connect psychological— here, sensory and perceptual—quantities on the one hand and physical quantities on the other. Examples are relations between visual brightness and light energy, loudness and pulse duration, warmth and areal extent of stimulation, and so on. Basically, the derivation of a psychophysical relation requires the correlation of two scales—one sensory, the other physical. Since the basic issue of the present paper is the interpretation of sensory scales, the role of psychophysical relations is central and crucial.

Psychosensory relations connect psychological (sensory) attributes of sensation alone. An example of a psychosensory relation is the statement that the loudness of a sound heard by two ears equals the simple sum of the loudnesses heard by each ear individually. We could write $L_b = L_l + L_r$, where L_b is binaural loudness, and L_l and L_r left-ear and right-ear loudnesses, respectively. This equation contains no quantitative reference to the physical stimuli that produce the loudnesses. Though one's initial reaction might be that psychosensory laws can be independent of psychophysical laws, this is not the case. In some instances, psychophysical and psychosensory relations appear capable of determining one another.
psychosensory relations appear capable of determining one another.

Finally, let us consider sensory-physical relations. These consist of statements about interrelations among physical dimensions alone. An example is Bloch's law of temporal summation, which states that in order to produce a constant visual effect, e.g., constant brightness, the product of flash duration (t) and luminance (L) must be constant ($L \cdot t =$ constant). This sensory-physical relation can readily be determined without knowing anything quantitative about the psychophysical relation between brightness and $L \cdot t$. Over the years, determination of sensory-physical relations, typically by direct matching, has formed the substantial base of most work on sensory processes, and it has been fruitful in yielding important knowledge about how the senses operate—this without having to invoke sensory scales.

But the relation between psychophysics and sensory physics may not be so remote as it seems at first blush. There are two types of interrelation. First of all, scales of sensation must be consistent with appropriate sensory-physical measurement. An analogy to

physical measurement may make this point clear. Assume that two objects—e.g., a lump of coal and a diamond—exactly balance when set on balance pans, i.e., have equal mass. A minimal requirement to be met by any system for scaling mass is that it assign the same scale value to the diamond and to the coal. If the scale values assigned are not the same, the system does not provide a valid scale of mass (though it may provide a valid scale of something else, such as volume or monetary value). Similarly, a system for scaling brightness or loudness (psychophysics) must assign to sensations that are equally bright or loud (sensory physics) scale values that are equal. This is the first type of interrelation. The second stems from situations where results of sensory-physical measurement restrict possible numerical representations (scales) even further. As the discussion later of visual processes makes clear, sensory physics can be more consistent with some scales than with others.

TWO TYPES OF SENSORY SCALE

For every dimension of sensation, there seem to exist at least two meaningful quantitative scales. Both scales are real and important, but usually they are different. As has been said, Type I scales are (roughly) approximated by scales derived from ratio-scaling procedures, Type II scales by scales derived from interval-scaling procedures. This point of view contrasts sharply with the positions of some investigators, like Stevens (1971), Anderson (1972), and Curtis, Attneave, and Harrington (1968), that only one of the types is valid, the other derivative and in some way not valid. The two types of scale reflect different modes of perception.

The two different modes of perception pervade the sensory domain, e.g., in the perception of loudness and brightness. When a subject reports Sound A to have loudness $= 1$, Sound B to have loudness $= 2$, and Sound C to have loudness $= 3$, one might conclude that the difference between A and B equals the difference between B and C. But if you then ask him whether these differences appear equal, he responds negatively. What the subject has done is transfer his mode of judgment from one aspect of his sensations to another, from the perception of sensory intensities to the perception of sensory dissimilarities. B sounds more like C than like A.

Formulation of the Theory

The following two equations express the core of the present theory:

$$\psi_I = f(\varphi) \tag{1}$$

$$\psi_{II} = g(\varphi_a, \varphi_b) \tag{2}$$

The functions f and g are psychophysical functions, Types I and II, respectively. On intensive continua, like brightness and loudness, f and g are often power functions, to wit

$$f(\varphi) = k\varphi^\beta \tag{3}$$

$$g(\varphi_a, \varphi_b) = c(\varphi_a{}^a - \varphi_b{}^a) \tag{4}$$

Thus, given physical stimulus of intensity φ, subjective magnitude (ψ_I) is a power function (exponent β) of φ, whereas, given stimuli of intensities φ_a and φ_b, the subjective dissimilarity (ψ_{II}) is the difference between each intensity raised to a power (exponent a). Empirically, as we shall see, a frequently turns out to be about half the size of β.

To repeat, ψ_I is interpreted here as a *sensory intensity*, ψ_{II} as a *sensory dissimilarity or difference*. The two types of scale defined by Eqs. 1-4 deal with different aspects of sensory intensity, but of course intensity is itself only one of the dimensions of sensory experience. We presume there to be additional psychological scales appropriate to all the other dimensions, for example, to sensory quality. Power functions may not, however, apply to the other psychophysical functions. And even though the present paper will restrict its concern to sensory magnitudes and dissimilarities that pertain to the intensive aspect of sensations, it is useful to bear in mind that these other, qualitative dimensions also exist. In particular, the Type II scales discussed here probably represent segments of more complex structures, specifically, projections of distances along the intensive dimension of a multidimensional similarity space.

The Approach via Sensory Physics: Scales for Brightness

Given these assertions, the first order of business is to try to demonstrate that there are indeed two different types of scale, both of which are meaningful. In the present section, we consider how the quantitative form of the Type I function for brightness magnitude relates to temporal processing of light intensity. To do this, we rely on theory of brightness as it applies to results of experiments that themselves need not employ scaling procedures at all.

Let us begin with some observations made by Fuortes and Hodgkin (1964) on electrical responses of the eye of the horseshoe crab, *Limulus*. Fuortes and Hodgkin examined the magnitude and the time course of the changes in receptor potential produced by brief flashes of light. When light energy was weak, the visual cells responded on a basically linear fashion, i.e., like a linear filter system. The magnitude of the output was proportional to the input intensity, and its temporal course had an exponential

characteristic. When light energy was stronger, however, the behavior of the cells became markedly nonlinear, with respect both to response magnitude and to response speed. Changes in receptor potential were no longer proportional to input energy, and, furthermore, the changes occurred more and more rapidly the greater the level of energy.

The quantitative results that emerged from Fuortes and Hodgkin's study were striking and of great theoretical import. It turned out that the changes in speed of the cellular responses were directly proportional to the magnitude of the output itself. The authors postulated a relatively simple model to account for that outcome: the nonlinear behavior of the system arises from the action of a feedback mechanism; the final output feeds back to earlier stages and serves both to diminish the system's overall gain and to increase its speed.

The observations and hypotheses recorded by Fuortes and Hodgkin were noted by Matin (1968) and by Sperling and Sondhi (1968), who incorporated feedback mechanisms into models that purport to account for human visual sensory physics, in particular for thresholds of luminance and flicker detection. At the core of these models is the notion of a visual time constant whose size depends on the magnitude of the system's output, and thus varies with the energy of the input. That notion is central to these models because measurements of luminance discrimination show that the critical duration of temporal summation—the limit on the time over which energy is integrated—depends systematically on the luminance of the background or comparison light against which test stimuli are presented (e.g., Graham & Kemp, 1938; Keller, 1941).

Most important to the present argument is the simple proportionality between response speed and response magnitude, where "response" is interpreted as visual intensity (brightness). The model for brightness has already been described in detail (Marks, 1972) and thus will only be summarized here. It begins with the postulation of a two-stage feedback system in which output from the second stage modulates sensitivity of both stages. Sperling and Sondhi (1968) came to conclude—on the basis of a survey of data on flicker and luminance discrimination—that the feedback portion of the visual system does contain two stages. If the input to such a two-stage system is a long pulse of light, the magnitude of the output will approximate the cube root of the intensity of the input. [In general, if there are n stages in the feedback system, the output will grow as the $1/(n + 1)$ power of the input.]

An important feature of brightness vision is its temporal characteristic. After a light is turned on, brightness first increases until it reaches a maximum, after which brightness declines a little, then remains roughly constant. The maximum in brightness is often called the Broca-Sulzer (1902) "enhancement."

Importantly, the speed of the brightness response (the reciprocal of the time required to reach maximum brightness) varies directly with luminance: the greater the luminance, the faster the rise to maximum. J. C. Stevens and Hall (1966) suggested, on the basis of their magnitude-estimation data, that speed is proportional to the cube root of luminance, an idea picked up by Stevens (1966) and by Anglin and Mansfield (1968). The most extensive empirical evidence has been provided by Mansfield (1973), who determined the temporal locus of the Broca-Sulzer maximum under many conditions of stimulation. The subject's task was simplicity itself: to vary duration so as to maximize brightness. So long as the test field is not too small, speed of brightness response (reciprocal of time to reach maximum) is proportional to the one-third power of luminance. Just as the two-stage model predicts a cube-root psychophysical relation between brightness and the luminance of long flashes of light, so too it predicts a cube-root sensory-physical relation between response speed and luminance.

The two-stage feedback model makes a second prediction about visual psychophysics, namely that the relation between brightness and luminance will depend on stimulus duration. As was already stated, given a long-duration pulse of energy, the input-output function of a two-stage system will follow a cube-root relation. If the input is a very brief pulse, the function becomes square-root. Both brightness matching (Aiba & Stevens, 1964; Katz, 1964; Nachmias & Steinman, 1965) and direct estimation of brightness (Raab, 1962; J. C. Stevens & Hall, 1966) show that brightness grows more rapidly with luminance at short durations than at long, and, quantitatively, the results are compatible with a ratio of exponents of 0.5:0.33.

Let us tie this all together. A simple model of brightness vision—which postulates two stages of processing that behave like low-pass filters with feedback—makes a number of predictions about both visual psychophysics and sensory physics. First, given a pulse of constant stimulus intensity, the model predicts that over time brightness increases (temporal summation), passes through a maximum, then declines to a steady state. Second, the model predicts that the speed of response (reciprocal of time to reach maximum) is proportional to the cube root of luminance. Third, it predicts the magnitude of the brightness response to be proportional to the cube root of luminance when duration is long, but to the square root when duration is short.

Brightness matches demonstrate the veracity of the first two predictions and are consistent with the third. Psychophysical judgments—for instance, magnitude estimates of brightness—suggest the veracity of all three predictions. Our special concern is the third one. Stevens and J. C. Stevens (1960) reviewed results of ratio-scaling experiments that involved fractionation, magnitude estimation, and magnitude

production; they concluded that scaled brightness of not-too-small, not-too-brief stimuli grows as the cube root of luminance. Note that few individual experiments gave this result. The variation in power-function exponent can be sizeable, both from experiment to experiment and from subject to subject within a given experiment. For instance, Marks and J. C. Stevens (1966) reported brightness functions for individual subjects, all of which were consistent with power relations, but which varied in exponent from 0.21 to 0.53, a range of 2.5:1. Nevertheless, the *average* exponent came out to 0.33.

Even should the model just described turn out not to be fully correct, there is good reason to assume direct proportionality between response speed and response magnitude, and thus to conclude that a cube-root function relates brightness to the luminance of long flashes of light. The most parsimonious hypothesis is then that *on the average* the ratio-scaling procedures yield results that closely approximate the underlying Type I scales of brightness.

The Approach via Scaling and Psychosensory Laws: Scales for Loudness

Type I Loudness Scales. Does the same sort of generalization hold for loudness, i.e., do ratio-scaling procedures, on the average, approximate the underlying Type I scale? Figure 1 and Table 1 give a median power-function exponent of 0.54. Stevens (1972) recommended a value of 0.67; this recommendation was based to a large measure on results of cross-modality matching experiments. Recently, Cross (1973) showed that when sequential biases are eliminated from magnitude estimates, the exponent falls close to 0.67. The present section shows that consideration of psychosensory relations also suggests that the Type I exponent for loudness falls in the range of 0.5-0.67.

In a classic study that is now four decades old, but amazingly modern in several respects, Fletcher and Munson (1933) tried to derive a general system and scale of loudness. A basic feature of their approach was the attempt to generate a single, consistent scale of loudness, not through the use of any subjective scaling procedures, but instead through consideration of auditory processes. In particular, Fletcher and Munson utilized the addition of individual loudnesses in multicomponent sounds and the addition of loudnesses by the two ears. The basic data employed were loudness matches between single tones and tonal complexes and between tones heard by one and by two ears. On the *assumption* that the rule of loudness addition—both for multicomponent sounds and for binaural listening—is simple linear summing, Fletcher and Munson derived a single scale of loudness that was consistent with the loudness matches.

Figure 2 shows how the Fletcher-Munson scale of loudness plots against loudness level (sound-pressure

level of an equally loud tone at 1,000 Hz). Over much of its range, the Fletcher-Munson scale is close to a 0.67 power of sound pressure. Of course, the validity of this scale is predicated on the validity of the linear rule of loudness addition. Fletcher and Munson had no means to test linearity.

The critical experiment to decide the issue applied the theory of conjoint measurement to ordinal judgments of loudness (Levelt, Riemersma, & Bunt, 1972). Conjoint measurement theory may be applied to situations where two or more dimensions of stimulation are varied; the theory permits one to determine whether the matrix of ordinal relations— here, greater or less loudness—is consistent with linearly additive effects over the two or more dimensions (Luce & Tukey, 1964).

Levelt et al examined the way that the two ears add loudnesses. The subects listened to dichotic stimuli—1,000-Hz tones that varied in sound pressure. Thus any given stimulus consisted of one sound pressure presented to the left ear, another sound pressure presented to the right ear. Given different pairs of such stimuli, the subject's task was simply to judge which stimulus appeared the louder. The results of the experiment were consistent with the psychosensory rule of simple additivity of loudnesses by the two ears, as Fletcher and Munson had
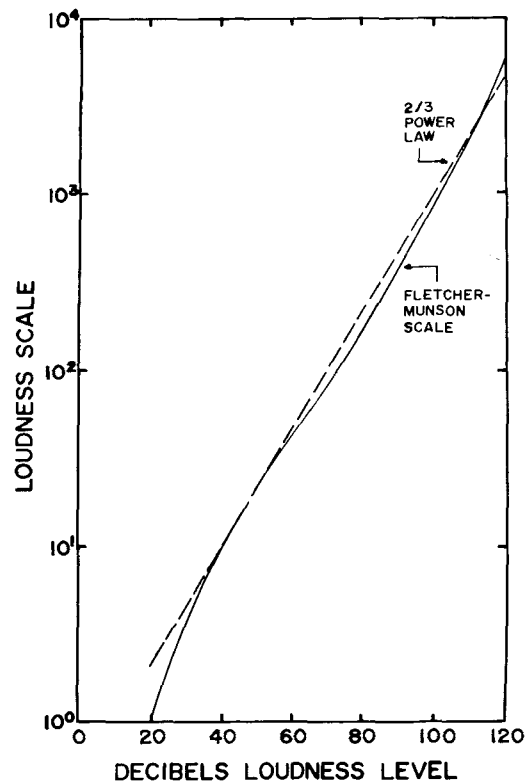

Fig. 2. Loudness scale derived by Fletcher and Munson (1933), plotted against loudness level in decibels. Shown also is a 0.67 power function of sound pressure.

hypothesized; that is, total loudness is the linear sum of the individual, component loudnesses. Conjoint measurement theory also provided a means whereby scales of loudness could be constructed from the ordinal (nonmetric) data; these scales could be described as power functions of sound pressure, with an average exponent of about 0.5. The exponents derived from scales of Fletcher and Munson and of Levelt et al resemble the exponents obtained from ratio scaling.

Type II Loudness Scales. Turning again to the compilation of power-function exponents given in Table 1 and Fig. 1, recall that the median exponent derived from ratio-scaling experiments is 0.54, that from interval-scaling experiments, 0.27. The results from ratio-scaling experiments appear roughly consistent with data on loudness summation. Do results from interval-scaling experiments also approximate some meaningful, underlying scale? The majority of the studies listed in the second part of Table 1 employed variants of the procedure called category rating, in which the subject attempts to use integer numbers to mark off constant sensory distances. In category production, the subject adjusts stimulus intensity as a response and the experimenter gives the numbers as stimuli. Procedures of bisection and equisection call on the subjects to adjust stimulus intensity or intensities in order to mark off equal sense intervals directly. But studies that are most important to our immediate purpose are those that involve direct estimation of sensory intervals. Here the subject is presented pairs of stimuli, one pair at a time, and asked to estimate numerically the subjective difference between the members of each pair.

The main reason why interval estimation is especially important is that the data contain a strongly determined metric structure. Not only is it possible to determine power-function exponents, and, therefore, numerical scales, directly from the numerical estimates, but it is also possible to obtain numerical scales from the rank orders of the judgments alone. This latter procedure of nonmetric analysis derives from the work of Kruskal and Shepard (e.g., Shepard, 1966). Interestingly and importantly, the underlying numerical scales obtained by the two procedures of data analysis (metric and nonmetric) are collinear (Rule, Curtis, & Markley, 1970).

Beck and Shaw (1967) were the first to report estimates of loudness intervals. They noted that the average estimates could be predicted fairly well by differences in loudness calculated from Garner's (1954) lambda scale. The lambda scale was itself determined primarily by an equisection procedure, and, as Table 1 shows, loudness in lambda units grows as the 0.26 power of sound pressure. In good agreement with this, analysis of Beck and Shaw's results yields power-function exponents of 0.3. More recently, Parker and Schneider (1974) obtained estimates of loudness intervals and of loudness

similarities, to which they applied nonmetric analysis. The scales they derived were even more like the lambda scale; those scales were describable by a power function with an average exponent of 0.26.

[Exponents obtained from interval estimation of loudness and reported by Dawson (1971) are quite a bit larger, however, than 0.26, as Table 1 shows. The reason for the discrepancy is not clear. Perhaps Type II psychophysical relations are intrinsically even more variable and susceptible to procedural effects than are Type I relations.]

Thus it turns out that nonmetric methods can yield two rather different scales of loudness. One scale arises when subjects judge loudness itself, the other arises when they judge loudness intervals. Both scales relate to sound pressure by power functions, but with exponents that differ by about a factor on the order of two. It is of some interest to note, in this regard, Garner's (1959) application of his lambda scale to the question of loudness additivity by the two ears. Garner was induced to conclude that additivity is not linear, but instead root-sum-square. Total lambda loudness equals the square root of the sum of the squares of the component lambda loudnesses. This conclusion contrasts with the near-linear additivity implied both by results of conjoint measurement and by data obtained with ratio-scaling procedures like magnitude estimation (see Marks, 1974). A similar contrast appears with respect to the rules of additivity of different tonal components presented to a single ear. Given that the tones are well enough separated in frequency to preclude mutual masking, the lambda scale predicts nonlinear additivity, whereas average results of ratio scaling predict linear additivity. Put another way, given loudness matching data for binaural listening and for complex tones, it follows directly from linear additivity of loudness components that the psychophysical relation approximates a power function with exponent of 0.5-0.6.

The Approach via Internal Consistency

Basic to proving the validity of any set of scales—Type I or Type II—is demonstrating their internal consistency. Consider the domain of pure tones that vary with respect to both frequency and sound pressure. A fundamental sensory-physical operation in the study of audition is cross-frequency loudness matching. That operation entails presenting to a subject tones of different frequencies; the task is to adjust the sound pressures such that the loudnesses are equal. The minimum constraint we must place on any scale of loudness is that it always assign the same number to the same loudness, regardless of how that loudness is brought about. When several different tones appear equally loud, they must be assigned the same scale value.

One way to assure internal consistency is to incorporate the results of direct matching with results of scaling. Thus we might begin with the loudness

scales determined by Levelt et al (1972) for 1,000-Hz tones and use cross-frequency loudness matches in order to derive an internally consistent scheme of scales for other sound frequencies.

Alternatively, we may look at the various direct scaling procedures themselves and inquire whether results obtained when the procedures are applied provide internal consistency. Lest it be thought that this criterion is trivial, it is worthwhile mentioning in advance that some scaling procedures, at least under certain conditions, fail to yield the required consistency.

Intramodal Consistency: Magnitude Estimation. Fortunately, scales derived from ratio-scaling procedures do pass the test of internal consistency. Let us consider, as one example, Hellman and Zwislocki's (1964) study of how noise masks the loudness of a 1,000-Hz tone; their procedures included both ratio scaling (magnitude estimation and magnitude production) and direct matches between masked and unmasked tones. Agreement between psychophysics and sensory physics was quite satisfactory. Internal consistency is also evident in schemes like Stevens's (1972) method for calculating the loudness of complex sounds, which permits precise computation of matches between pure tones and any complex sound spectra.

Intramodal Consistency: Category Rating. Interestingly, at least two studies appear to show that scales derived from rating procedures can violate consistency. Both experiments required subjects to rate the brightnesses of flashes of light that varied with respect to duration. It may be recalled from our earlier discussion that the brightness produced by a light of constant luminance increases as its duration grows, until it reaches a maximum value, but that the time needed to reach maximal brightness itself decreases as luminance increases. These basic facts have been demonstrated both by direct interocular brightness matching and by magnitude estimation, again demonstrating the internal consistency of the latter procedure.

Raab, Fehrer, and Hershenson (1961) asked subjects to rate the brightnesses of flashes of light on a 9-point scale; only a single luminance was employed, but duration varied from 10 to 500 msec. With the level of luminance employed (3,000 fL), brightness is known to reach its Broca-Sulzer maximum at a duration of about 10 msec. In fact, the category ratings suggested no maximum, but rather a steady increase in brightness over the period of time examined. A more elaborate version of the same experiment was conducted by Lewis (1965), who presented six levels of luminance, 10 durations, and used an 11-point scale. Not only did the ratings show no Broca-Sulzer maximum, but neither was there any indication of the systematic change with luminance in the critical duration of temporal summation. These results, obtained by category ratings, are, therefore,

at marked variance with the known facts of brightness vision.

The obvious conclusion to be drawn is that rating scales do not provide valid measures of sensory magnitudes, since they do not provide internal consistency. But even though this conclusion may be correct, it still may be possible, by interpreting the results of Raab et al and of Lewis in a slightly different manner, to salvage the view that rating scales can provide valid measures of something. Note that those results imply that ratings are inconsistent with *brightness*, i.e., inconsistent with the psychological dimension of visual intensity whose scale is of Type I. But, as we have said, ratings do appear proper and applicable to the derivation of scales of Type II. Therefore, it may be appropriate to consider the ratings obtained by Raab et al and by Lewis as measures of *brightness dissimilarity*. If we are willing to assume that the sensory physics of brightness dissimilarity may not be the same as the sensory physics of brightness per se, then category ratings may turn out to be consistent after all, consistent with brightness dissimilarity but not with brightness. On the other hand, it may well be (and probably is) that the sensory physics of brightness and of brightness dissimilarity really are the same. If so, then it would be necessary to examine in detail the procedure of category rating in order to determine whether any of its methodological versions can prove to be an internally consistent measure of sensation.

Cross-Modal Consistency: Cross-Modality Matching. The method of cross-modality matching requires subjects to adjust stimulus intensities on Continuum 1 so that the sensations they produce "match" those sensations produced by stimuli on Continuum 2. Although use of the method itself dates back to the 19th century (e.g., Münsterberg, 1890), its maor development arose from attempts to determine whether ratio-scaling procedures like magnitude estimation provide consistency. If the two Type I power equations

$$\psi_{I_1} = k_1 \varphi_1{}^{\beta 1} \qquad (5)$$

$$\psi_{I_2} = k_2 \varphi_2{}^{\beta 2} \qquad (6)$$

describe results obtained from scaling sensory intensity on Continua 1 and 2, respectively, then cross-modality matches between ψ_{I_1} and ψ_{I_2} predict the equation between pairs of physical stimuli

$$\varphi_1 = a\varphi_2{}^{\beta 2/\beta 1} \qquad (7)$$

where $a = (k_2/k_1)^{1/\beta 1}$. Thus cross-modality matches should be describable by a power function whose exponent equals the ratio of the exponents of the psychophysical functions. That result has been verified for a very large number of comparisons, including loudness, shock, and vibration (Stevens,

1959), force of handgrip and heaviness (J. C. Stevens & Mack, 1959), handgrip and shock, heaviness, pressure, vibration, loudness, and brightness (J. C. Stevens, Mack, & Stevens, 1960), and brightness and loudness (J. C. Stevens & Marks, 1965).

The cross-modality matches between perceived length of lines and loudness and between perceived length and brightness (Stevens & Guirao, 1963) are of particular importance from a theoretical point of view. Perceived length of lines is very nearly proportional to actual physical length, i.e., the exponent of the Type I power function is close to unity. As we have already seen, the size of the Type I power-function exponent for loudness is about 0.6, that for brightness, about 0.33 (when neither size nor duration is very small). The exponent obtained from cross-modality matching of length to loudness was about 0.7, that from matching length to brightness about 0.3. Though the former value is a little high, nonetheless the agreement with prediction is reasonably good.

Cross-modality matching demonstrates the internal consistency (transitivity of power-function exponents) of psychophysical functions of Type I. It appears eminently reasonable that, in such a system of functions, the exponent governing perceived length approximate unity. A similar view was expressed by Krantz (1972), who pointed out that the transitivities demonstrated by cross-modality matching yield as a minimum the ratios of exponents applicable to various modalities. If, then, the exponent for length is taken to be 1.0, it becomes possible to calculate absolute exponents for the other sensory and perceptual continua. If we continue on this approach, the results reported by Stevens and Guirao (1963) imply Type I exponents equal to about 0.7 for loudness, 0.3 for brightness.

Relation Between Type I and Type II Scales

Another question we may ask is whether there exists consistency (cross-modal transitivity) for psychophysical functions of Type II. For example, do the exponents a of Eq. 4 predict cross-modality matches? To the extent that the ratio of a to β remains constant across modalities, Type II psychophysical functions will prove to be just as consistent cross-modally as are Type I functions. On the other hand, it may be that the observed consistency with Type II functions is

fortuitous, since the relevant cross-modal comparisons should involve sensory differences or dissimilarities, not sensory magnitudes.

Stevens's (1971) survey of ratio- and interval-scaling experiments on loudness suggested that the Type II exponent, a, is about one-half the Type I exponent, β (cf. the present Table 1). Stevens termed the exponent derived from interval-scaling procedures as "virtual," in contrast, presumably, to the "actual" exponent derived from ratio-scaling procedures. This dichotomy in terms seems to express Stevens's belief that only ratio-scaling procedures can yield "valid" scales of sensation.

Table 2 lists average or prototypical values of Type I and Type II exponents of power functions for five sensory and perceptual modalities. In every example, except perhaps for the ambiguous results on visually perceived area, the ratio $a:\beta$ approximates 0.5. Thus there is a strong possibility that a constant relation exists between Type I and Type II scales, whereby sensory distances are proportional to differences between the square roots of sensory magnitudes

$$\psi_{II} = b(\psi_{Ia}^{1/2} - \psi_{Ib}^{1/2}) \tag{8}$$

Such a generalized relation between Type I and Type II scales would add to the likelihood that one of the modes of perception is derived from the other. On purely logical grounds, the transformation could go either way: Type I scales primary and Type II derivative, or vice versa. There is some reason, though, to suspect that generation of magnitudes precedes generation of differences.

Some light can be shed on this issue by considering scales of loudness. As we have already seen, loudness addition (in complexes and by the two ears) is linear at the level of Type I perception but square root of sum of squares at the level of Type II perception. A simple account is provided by the hypothesis that sound stimuli are mapped into Type I loudness (primarily by power functions, exponent approximately 0.6), after which component loudnesses add linearly to form total loudness. Total loudnesses (Type I) are then transformed by a square-root function and subtraction into loudness differences (Type II). Difficulty is encountered, however, if we assume the reverse order of transformations (stimulus to Type II

Table 2
Exponents β of Type I and a of Type II Psychophysical Functions

Continuum	β	Source	a	Source
Loudness	0.6	Present Survey	0.27	Present Survey
Brightness	0.33	Stevens & J. C. Stevens, 1960	0.16	Curtis, 1970
		Marks & J. C. Stevens, 1966		
Heaviness	1.4	Stevens & Galanter, 1957	0.6	Curtis et al, 1968
			0.7	Marks & Cain, 1972
Perceived Area	0.7	Teghtsoonian, 1965	0.7	Rule et al, 1970
			0.4	Marks & Cain, 1972
Roughness	1.5	Stevens & Harris, 1962	0.7	Marks & Cain, 1972

to Type I). Recall that additivity of components takes place at the level of Type I perception. If Type I loudnesses were derived from Type II loudness differences, then at the level of Type II there would not yet be total loudness. Thus Type II loudness could not be consistent with the sensory physics of loudness addition. In order for both Type I and Type II scales to be consistent with sensory physics, Type I must precede Type II.

What purpose might be served by a transformation from Type I to Type II? One possibility arises from the interpretation of Type II scales as scales of dissimilarity. Overall dissimilarity between stimuli is, in any given modality, usually compounded of variations along several perceptual dimensions: intensity, quality, etc. Transformation from Type I to Type II might serve, by constricting the range of variation along any one dimension, to prevent total range of dissimilarity from getting too large when perceptions vary along several dimensions. The square-root nature of the transformation could be the outcome of a process whereby perceptual magnitude is normalized by its variability. This view is related to the hypotheses of Stevens and Galanter (1957), Eisler (1962), Stevens and Guirao (1962), and others, to wit, that discriminability or variability is responsible for the oft-noted nonlinear relation between category and magnitude-estimation scales.

SOME IMPLICATIONS AND APPLICATIONS

Rating Scales

In light of the present theory, we interpret the task of stimulus rating as one in which subjects attempt to place stimuli into categories that reflect constant steps of sensory dissimilarity (cf. Ekman, Goude, & Waern, 1961). Now, all scaling procedures are susceptible to effects of variations in procedure; rating scales are, in particular, notoriously sensitive to variables such as context (Stevens & Galanter, 1957; Parducci, 1965). Frequently, the influence of variables like stimulus range and number of available categories manifests itself as a change in the size of the exponent of a fitted power function (Marks, 1968). To some extent, these effects may be an outcome of the tendency on the part of subjects to use all rating categories equally often. Stevens and Galanter (1957) pointed out that "bias" induced by such a tendency may be eliminated, or at least minimized, by iterative experimentation, i.e., by adjusting the spacing among stimuli from one experiment to the next until the responses become symmetrically distributed (see also Pollack, 1965a, b). Stevens and Galanter conducted a short-cut approximation that yielded a "pure" rating scale for loudness. As can be seen in Table 1, their "pure" scale may be described as the 0.26 power of sound pressure (Marks, 1968); that is, it is the same scale as that obtained by equisection and by nonmetric analysis of judgments of loudness difference and of loudness dissimilarity. Thus it appears that category ratings can, under appropriate conditions, yield good estimates of Type II sensory scales.

The present interpretation of rating scales as interval (Type II) scales is based more on empirical outcome than on theory. There is at present no compelling theoretical reason why ratings should reflect dissimilarity rather than magnitude. But, in general, ratings more closely approximate Type II than Type I scales. An exception is Gibson and Tomko's (1972) report that, under an appropriate experimental condition, rating scales and magnitude-estimation scales were collinear. The appropriate condition consisted of selecting the end points of the rating scale to coincide with the previously determined range of magnitude estimates. Thus it may be that, given special choices of procedure, category ratings can approximate Type I scales as well as magnitude estimates can.

Thus, at least two possible interpretations of rating scales exist. One states that ratings are not intrinsically approximations solely to Type I or to Type II scales; they may approximate either scale, depending on instructions and procedure. The other interpretation states that ratings are, fundamentally, approximations to Type II (interval) scales. Under this interpretation, variations in parameters like stimulus spacing may serve to "bias" the rating scales either in the direction of Type I scales or away from them.

Functional Measurement

There is a second way to use rating procedures in order to get at "unbiased" scales of Type II, namely through the approach known as functional measurement. Functional measurement is Anderson's (e.g., 1970) term for the simultaneous determination of sensory scales and psychosensory relations (as the latter have been called here). In its essence, it applies to experiments in which several stimulus variables may be assumed to contribute separate effects, the concatenation of effects being either additive or multiplicative. Often it is possible to determine simultaneously both the model of concatenation that is appropriate and also the underlying sensory scales. The application of conjoint measurement to the question of loudness additivity is very much akin to functional measurement.

Several of the studies conducted by Anderson and his colleagues have employed tasks that require the subects to "integrate information," for example, to judge the "average heaviness" (Anderson, 1972) or the "average grayness" (Weiss, 1972) of pairs of stimuli. When the judgments were made on rating scales (numerical and graphical), the results of both experiments were consistent with simple models of linear averaging of subjective values (linear psychosensory laws). Concomitantly, the psycho-physical relations between sensation and stimulus

appear to be describable by power functions; the exponents are somewhat smaller than unity both for heaviness and for grayness.

By way of contrast, simple models of linear averaging were not applicable when the response procedure was magnitude estimation. Further, the power-function exponents obtained from magnitude estimation are also different, larger than those obtained from ratings. Although Anderson has interpreted these outcomes to mean that category rating is basically valid, whereas magnitude estimation is basically invalid, there exists an alternative interpretation that is more consistent with data already considered. Under the present interpretation, the averaging task is appropriate to Type II perception, in that subjects interpret "average heaviness" or "average grayness" as "equisimilarity." Thus, application of functional measurement to ratings yields good estimates of Type II scales. The failure of linear additivity to apply when the procedure employs magnitude estimation, however, does not necessarily stem from any basic lack of validity of magnitude estimation; instead, the failure may stem from the inappropriateness of the linear model. Although there is no a priori reason why this should be so, nevertheless it appears that "averaging" has much in common with "bisecting."

There are some data that create an apparent dilemma for the present interpretation—and for other formulations, such as Anderson's. When the subject's task is to judge "average length" or "average inclination" of a line, it turns out that magnitude estimation does yield data that are consistent with linear averaging (and also consistent with linear psychophysical functions) (Miller & Sheldon, 1969). Additional evidence was provided by Weiss and Anderson (1969), who also showed that judgments of "average length," as determined by a production procedure, are consistent both with linear averaging and linear psychophysical functions. And judgments of similarity—which usually tap Type II perception— also yield linear psychophysical functions for length (Young, 1970).

It may be that perceived length is exceptional, in that its Type I and II scales are identical.[1] Length is, of course, an extensive, rather than an intensive, continuum. It seems likely that most people could manipulate mental images of line segments in order to produce judgments proportional to physical extent.

In any case, since magnitude estimation of length typically yields linear psychophysical functions, these results imply that the procedure of magnitude estimation is not necessarily invalid and that the task of "averaging" can be appropriate to magnitude estimation when the continuum is perceived length. We could assume that when other continua are used, it is not averaging, but magnitude estimation that fails. It seems more likely, however, that the reverse is true: magnitude estimation is appropriate, but

averaging is not. Again, it may be pointed out, a linear psychophysical function for perceived length is consistent, in the context of cross-modality matching, only with the Type I psychophysical functions determined by ratio-scaling procedures such as magnitude estimation and production.

Scales of "Subjective Number"

J. C. Stevens (reported by Marks, 1968) obtained category ratings using numbers themselves as stimuli. It turned out that the power-function exponent for ratings of numbers equalled 0.54. Attneave (1962) estimated the exponent to be 0.4, on the basis of subjects' bisections of the interval 1-1,000,000. Since the system of Type I psychophysical functions has been formulated such that β for "subjective number" equals 1.0, the ratio of a to β again falls close to 0.5. An analogous, and probably related, result was obtained by Banks and Hill (in press), who devised the novel procedure of asking subjects to emit "randomly spaced" numbers. The spacing among numbers that resulted was nonlinear, and it suggested a power relation with exponent of about 0.6. A simple explanation is that the subjects produced numbers in such a manner that they were spaced to give equal steps of similarity. The same explanation applies also to Banks's (1973a) results obtained on the continua of force, areal extent, and electric shock. The exponents derived from random production are small compared to those obtained from magnitude estimation, because the two procedures tap different perceptual processes.

The nonlinear relation between "number dissimilarity" and objective number can also help us to understand the results of a recent study reported by Rule and Curtis (1973). They presented subjects with pairs of stimuli: one component of each pair was one of nine weights to be lifted; the other was one of the integer numbers from 1 to 9. The subject's task was to report which was relatively greater, the weight or the number. This novel procedure can be looked upon as a categorical (Type II) version of cross-modality matching between number and weight. Rule and Curtis applied the methods of conjoint measurement theory to the data, and they obtained psychophysical functions between weight and mass and between "subjective number" and actual number, both of which could be described as power functions. The exponents were 0.57 and 0.63, respectively. The latter value falls fairly close to the 0.54 determined from simple ratings of number. In addition, if we assume the ratio a/β to be approximately the same for both heaviness and number, then we would predict the scales that Rule and Curtis derived also to be consistent with the exponent obtained from magnitude estimation of heaviness (cross-modality matching of number and heaviness). That is what Rule and Curtis found. Those authors, however, interpreted their results in a different manner;

specifically, they argued that it is in magnitude estimation that a subjective scale of number is nonlinearly related to objective number. The present interpretation does not deny nonlinear mapping of number in magnitude estimation—the variation in loudness exponents (Table 1) attests to that. However, mapping of subjective to objective number is linear in Type I scales, which are very nearly averages of scales derived from magnitude estimation and magnitude production. Under the present interpretation, it is in category rating that the prototypical subjective scale of number is nonlinearly related to objective number.

Judgments of Sensory Intervals and Differences

As we have already seen, one way to approximate Type II scales is to ask subjects to judge sensory intervals, differences, or dissimilarities. The direct estimation of sensory intervals was established as an important procedure in the 1930s and 1940s, particularly as it was used in the scaling of surface colors. The renotation of the Munsell system of colors was based on a total of three million ratio judgments of color difference (Newhall, Nickerson, & Judd, 1943). One of the three dimensions of that system is Munsell value, which corresponds to the dimension lightness-darkness of surfaces. Saunderson and Milner (1944), Stevens and Galanter (1957), and Glasser, McKenney, Reilly, and Schnelle (1958) all demonstrated that Munsell value approximates the 0.33 power of reflectance. Power-function exponents derived from magnitude estimation of lightness of gray surfaces are clearly larger in size than that—they vary from 0.5 (Warren & Poulton, 1966) to 1.2 (Stevens & Galanter, 1957). Thus, again, the Type I exponent, β, appears to be quite a bit larger than the Type II exponent, a, although in the present case of lightness we have no firmly established value for β.

When we examine results of experiments in which subjects estimate sensory intervals, we come across a curious finding. Frequently, there appears to be a nonlinear relation between the derived intervals and the original numerical estimates of sensory intervals on which the derived intervals were based. Often the results obtained from the direct estimation of differences or intervals of sensory intensity can be described by an equation of the form

$$\psi_{int} = c(\varphi_a{}^a - \varphi_b{}^a)^\gamma \qquad (9)$$

(e.g., Curtis, Attneave, & Harrington, 1968). It should be apparent that Eq. 9 is like the Eq. 4 that governs Type II psychophysical functions, but with the addendum that the intervals are themselves raised to the power γ.

There are two points to be made about this result. The first is that whereas occasionally the exponent γ is

hardly different from unity (Curtis et al, 1968; Rule, Curtis, & Markley, 1970), more often γ deviates quite a bit from unity (Rule et al, 1970; Curtis & Rule, 1972). On the average, γ is greater than 1.0.

Curtis and his colleagues (Curtis et al, 1968; Rule et al, 1970) have argued that there is only one type of valid sensory scale, that the same scale is tapped both when the procedure is magnitude estimation and when it is interval estimation. The fact that power-function exponents obtained by magnitude estimation differ from those obtained from interval estimation is assumed to result from the nonlinear way that subjects use numbers (i.e., from the fact that $\gamma \neq 1.0$). In the terminology of the present paper, their hypothesis states that only Type II scales are valid, that the exponent, β, of the Type I power function is the product of the Type II exponent, a, multiplied by the "bias" exponent, γ. In agreement with that hypothesis, Curtis et al and Rule et al point to the empirical equality between the exponent obtained from magnitude estimation and the product of the two exponents obtained from interval estimation. But agreement between prediction and observation is not always very good (e.g., Curtis, 1970).

Given that the basic validity of Type I scales and psychophysical functions is established (by nonmetric procedures, by theoretical considerations of sensory physics, and by consideration of internal consistency) there must be some other explanation for the empirical evidence at hand. One possibility is that the task of interval estimation leads the subject not only to judge the dissimilarity (Type II relation) between each pair of stimuli, but also to apply Type II relations to his own numerical judgments. That is, the task may induce the subject to match number dissimilarity to sensory dissimilarity. If that be the case, then we would predict the results of interval estimation to conform to Eq. 9. The exponent γ becomes the reciprocal of the exponent a that applies to the Type II psychophysical function for "subjective number." The evidence considered in the last section points to an exponent a of about 0.5-0.6 for Type II subjective number. Thus γ should equal about 1.67-2.0. Reported values of γ range from 1.0 (Indow & Uchizono, 1960; Wright, 1965; Ramsay, 1968), through 1.1 (Curtis et al, 1968), 1.3 (Curtis, 1970), 1.5 (Curtis & Rule, 1972), to 1.7 (Indow & Kanazawa, 1960; Rule et al, 1970). That these estimates of γ are somewhat too low is not surprising, since they are based primarily on estimation procedures. Quite possibly, if production procedures were used, γ would turn out greater than 1.67-2.0. In any case, note that, if the ratio between Type I and Type II exponents for subjective number is the same as the ratio applicable to the modality judged, we would also expect the product of a and γ in Eq. 9 to equal the exponent β obtained from magnitude estimation.

Scaling of Multidimensional and Unidimensional Similarity

There has been a great deal of interest among students of perception in the nature of multidimensional perceptual space. Colored surfaces, for instance, vary along the psychological dimensions of hue, saturation, and lightness. The corresponding Munsell scales of hue, chroma, and value provide interval-scale measures of these dimensions, but the Munsell system provides no common unit. That is, a one-step difference in Munsell value is not necessarily equal, perceptually, to a one-step difference in hue or chroma. Two of the aims of multidimensional scaling of colors have been to verify the spacings given by the Munsell system and to determine a common unit of perceptual difference.

In multidimensional color space, the distances between points (objects) reflect their psychological difference or dissimilarity. As a matter of fact, it doesn't seem to matter whether instructions to the subject are to judge difference (Indow & Kanazawa, 1960; Indow & Uchizono, 1960; Ramsay, 1968) or to judge dissimilarity (Torgerson, 1958; Helm, 1964). In both cases, obtained scale values agree reasonably well with the Munsell notation. Thus, from the subject's point of view, difference equals dissimilarity.

The present paper has concerned itself with unidimensional scales of sensation and of sensation dissimilarity. The latter—the scales of Type II—may be considered collapsed, unidimensional versions of multidimensional similarity space. In its full form, similarity space contains several dimensions: three in the case of simple colors, two in the case of simple sounds (pure tones).

Psychological scales of tonal dimensions provide an intriguing and unusual puzzle. For, although the Type II scale of loudness dissimilarity differs clearly and markedly from the Type I scale of loudness, the Type II and Type I scales of pitch dissimilarity and pitch appear to be essentially identical! It is true that the revised mel scale of pitch (Stevens & Volkmann, 1940), which was based primarily on results obtained via an equisection procedure, does differ from the original mel scale (Stevens, Volkmann, & Newman, 1937), which was based on results obtained via fractionation. Nevertheless, that difference is relatively small, and, furthermore, both of those scales can be derived from the same procedure—magnitude estimation—given appropriate choice of location of the standard stimulus (Beck & Shaw, 1961). More notable is the invariance: for example, even category ratings can yield results approximating the mel scale (Stevens & Galanter, 1957). Parker and Schneider (1974) derived the mel scale by means of nonmetric analysis of estimates of pitch dissimilarities. Needless to point out, the fact that all of these different procedures yield the same underlying scale suggests that if any of the procedures can provide valid scales, all can, at least under certain conditions. An open question is how to interpret the

unitariness of the mel scale. Stevens and Galanter (1957) suggested that the mel scale might be only an interval scale, or, in the present terminology, a Type II scale. Thus it may be that there exists no Type I scale of pitch.

Before concluding this section, a few words should be said about the procedure of similarity estimation, which was first employed by Eisler and Ekman (1959) in the scaling of pitch. In the procedure of similarity estimation, the subject's task is to place the perceived similarity between two stimuli on a scale that goes from 0 (complete dissimilarity) to 100 (complete similarity or identity). That procedure has also been applied to the unidimensional continua of heaviness (Eisler, 1960) and grayness and perceived area (Ekman, Goude, & Waern, 1961), as well as to several sorts of multidimensional continua (Ekman, Engen, Künnapas, & Lindman, 1964). We would expect on the basis of the present theory that the scales that underlie similarity estimates be of Type II. However, the procedure of similarity estimation is itself somewhat peculiar; for, though the scale has a fixed zero point, and thus seemingly should for that reason have ratio properties, psychologically it cannot. From the subject's point of view, a similarity estimate of 50 should reflect a perception that is half as great as that given by 100. But 100 means identity! What identity really implies is zero dissimilarity.

For this reason, the metric properties of similarity estimates become questionable. It remains possible, of course, to analyze a set of similarity estimates nonmetrically; this I have done for the average estimates of heaviness reported by Eisler (1960) and for the estimates of grayness and area reported by Ekman et al (1961). Unfortunately, only five stimuli were used in the scaling of area, a number too small to yield, via nonmetric analysis, a well-constrained scale. Well-constrained scales could be derived for heaviness and for grayness. Heaviness similarity yielded a derived scale that related to mass by a power function with exponent of 0.26. That value appears rather low, e.g., in comparison to Type II scales of heaviness derived from interval estimation (Curtis et al, 1968; Marks & Cain, 1972) and from category ratings (e.g., Anderson, 1972). Grayness similarity, on the other hand, yielded a power-function exponent of 1.8, which appears to be quite high. Its large size may in part be due to simultaneous brightness contrast, but that explanation seems insufficient to account for so great a departure from 0.33. Thus it remains unclear to what extent scales derived from the procedure of similarity estimation agree with the Type II scales derived from category ratings and from estimates of intervals and dissimilarities.

Reaction Time

Time and again over the years there recurs the intriguing question of whether simple reaction time—the latency of response to the onset of a stimulus—can serve as a measure of sensation

intensity. I can recollect that question being discussed during my years as a graduate student—in the early to mid-1960s. Many of us who were working in sensory scaling came to wonder, at some point, if reaction speed was proportional to magnitude estimates of sensory intensity.

A strong affirmative statement of the relation between reaction time and sensation intensity was made by Piéron (1952), who tended toward the opinion that sensation itself cannot be measured. To wit: "If . . . a strict parallelism were admitted between reaction speed and sensation intensity, then the law governing the variation of the speeds should be taken to govern also that of the perceived intensities. What, therefore, is this law, derived, as a reciprocal function, from that which links the reaction latencies to the stimulus intensities? . . . The reducible part of the times is inversely proportional to the stimulus intensity raised to a certain power, which may be less, equal to, or greater than unity (Piéron's law, 1913) [Piéron, 1952, p. 352]."

As stimulus intensity increases, the reaction time decreases; it approaches an asymptotic value at high stimulus levels, a value called the irreducible minimum. The subtracted difference between reaction time and the irreducible minimum gives an adjusted reaction time whose reciprocal—reaction speed—Piéron presumed to parallel sensation intensity. His own early researches (Piéron, 1914) led him to propose that reaction speed relates to stimulus intensity by a power function.

Does reaction time, à la Piéron, mimic sensory intensity? Are the power-function exponents that relate reaction speed to stimulus intensity equal to the exponents that govern Type I psychophysical functions? Piéron's reaction-time data are not wholly satisfactory to answer the question, primarily because of the rather primitive methods of stimulation he was forced to employ. Fortunately, there exist some more modern experiments—in the realms of hearing, warmth, and vision—whose results may help point to an answer.

In an extensive investigation, Chocholle (1940) measured reaction times to auditory stimuli of different frequencies that varied over wide ranges of sound pressure (as much as 106 dB). If we take the results obtained at 500, 1,000, 2,000, 4,000, 6,000, and 10,000 Hz and assume an irreducible minimum reaction time of about 106 msec (about 4 msec less than the reaction times obtained at very high levels), we find that reaction speed approximates a power function of sound pressure with an exponent of about 0.3. That value is half the size of the Type I power-function exponent that governs loudness sensation; furthermore, as should be obvious, the reaction-time exponent falls quite close to the prototypical Type II exponent that governs loudness dissimilarity. [Luce and Green (1972) contended that auditory reaction speed is proportional to loudness—Type I in present terminology. To support

this contention, they replotted Chocholle's reaction-time data, minus an irreducible constant, and described the result as a power function with exponent -0.3 re sound intensity, i.e., equal to -0.6 re sound pressure. Inspection of Luce and Green's Fig. 7, however, shows that the exponent is -0.3 re sound pressure.]

A corroboratory result was reported by McGill (1961), who measured reaction times to 1,000-Hz tones and reported a power function with exponent also of -0.3. To McGill's question, "Are loudness and reaction time related by a square root law? [McGill, 1961, p. 199]," the answer appears to be "yes."

Were this result— that the exponent governing reaction time is about half the size of that governing sensory magnitude—limited to a single sense modality, we might consider it interesting, but possibly fortuitous—or at least peculiar to the auditory system. But recent evidence from another sensory domain—the warmth sense—also points to a square-root relation between reaction speed and sensory intensity. Banks (1973b) measured reaction times to radiant stimuli that varied with respect to both radiant intensity and areal extent of stimulation, and he derived functions that show how reaction speed depends on intensity. Each function could be described by a power equation whose exponent depended systematically on size: the greater the area, the smaller the exponent. In fact, the reaction-time functions mimic magnitude-estimation functions that were reported earlier (J. C. Stevens & Marks, 1971). The magnitude estimates were obtained with the same apparatus, same stimulus areas, and approximately the same stimulus intensities. The only notable difference between reaction time and magnitude estimation was in the sizes of the exponents: reaction time gave exponents consistently half as large.

To the best of my knowledge, vision is the only other modality for which there exists substantial reaction-time data that can be compared to scaling data. But here the results appear to throw a wrench into our scheme, for at first glance it seems that visual reaction speeds do agree with the Type I scale of brightness sensation. Reaction speed—like brightness—is proportional to the cube root of luminance (Vaughan, Costa, & Gilden, 1966; Mansfield, 1970). However, the simplicity of that outcome dissolves under more careful examination. Recall that Type I brightness grows as the one-third power of luminance only when flash duration is long, but as the one-half power when the duration is short. Reaction speed, on the other hand, grows as the one-third power of luminance regardless of flash duration. A related result is that reaction time displays neither a luminance-dependent change in critical duration of temporal summation nor a Broca-Sulzer maximum (Raab, Fehrer, & Hershenson, 1961; Vaughan et al, 1966; Mansfield, 1970). Thus visual reaction times are, to put it mildly, peculiar. Sounds that differ in frequency and pressure, but appear equal in loudness, give the same reaction time; radiant stimuli that differ

in area and intensity, but appear equally warm, give the same reaction time; lights that differ in duration and luminance, but appear equally bright, do *not* give the same reaction time.

Were it not for the enigmatic behavior of visual reaction times, we might be in a position to make a stronger conclusion. As it is, we must temper what seems an appealing hypothesis: that reaction speed—the reciprocal of reducible reaction time— is proportional to the square root of the intensity (Type I) of the sensation that is aroused.[2] A corollary may turn out to be that the reaction-speed exponent is the same as the exponent that governs sensory dissimilarity. In that case, equal increments in reaction speed would correspond to equal steps of sensory difference or dissimilarity. Perhaps this is why category ratings of brightness fail to display luminance-dependent critical duration or Broca-Sulzer maximum.

Animal Psychophysics

Recent years have seen a blossoming of interest in the sensory and perceptual responses of infrahumans —the realm of inquiry often called animal psychophysics. And among the diverse approaches and directions that comprise animal psychophysics is that of scaling. It is probable that the distinction between Type I and Type II scales of sensation applies as well to animals as to humans.

Not surprisingly, most of the attempts to measure the sensations of animals yield approximations to Type II scales. Herrnstein and van Sommers (1962), and later Boakes (1969), obtained bisections of brightness intervals from pigeons. The results were much like bisections obtained from humans (cf. Stevens, 1961b). Direct interval scaling of this sort is relatively easy to accomplish with infrahumans. In Boakes's experiment, for example, the pigeons were first trained to respond on one key to a high luminance, on a second key to a low luminance, then tested with intermediate luminances. The bisection point was taken to be that intermediate level of luminance that yielded equal proportions of response on the two keys. Operationally, this procedure is identical to the variation of the bisection procedure employed by Cross (1965) in his study of loudness bisections by humans.

It is possible, if not likely, that the direct interval-scaling procedures, such as bisection, will show the same influence of procedure when they are applied to animals that they show when they are applied to humans. If so, then the use of nonmetric procedures probably will be more productive in the determination of unbiased scales of Type II. Schneider (1972) required pigeons to discriminate whether various pairs of monochromatic lights were the same or different. Taking the percentage of correct responses as a measure of discriminability between lights, Schneider applied the Shepard-Kruskal method of analysis of proximities in order to generate a two-dimensional representation of color difference (Type II) in the pigeon.

The same difficulty— variations that affect the form of psychophysical functions—would also pertain to the determination of scales of Type I, except that it may be impossible to do direct ratio-scaling with animals. To data, no one has devised the equivalent to magnitude estimation. But, in any event, it is certainly possible to approach scales of Type I via psychosensory laws—that is, by applying nonmetric procedures (e.g., conjoint measurement). The question of how components add their loudnesses in tonal complexes is, in theory, as readily answerable for animals as it is for humans. A second approach is via sensory physics—that is, by theoretical considerations of sensory, but nonscaling, data. For instance, we expect that other organisms besides man will demonstrate reciprocality between speed and magnitude of the visual system's response.

There perhaps exists a third way to get at Type I scales of sensation in animals. It is only a "perhaps" because the approach is predicated on finding a constant and consistent relation between Type I and Type II scales. Given that the generation of unbiased scales of Type II would not prove difficult, then, if scales of Type I bear a constant relation to scales of Type II, the former may be derived from the latter.

Although the psychological study of animal sensation is fascinating in its own right, there is the bonus to be derived from scaling, namely the potential to correlate psychophysics with physiology. Once scales of sensation are generated with animals in a normal physiological state, one might wish to investigate the effects of various types of physiological intervention (neural excision, drugs, etc.) or even to record neural responses concurrent with psychophysical determinations. Results of such experiments could be of enormous benefit to the understanding of the interrelation between physiology and sensation.

What Kind of Measurement do Type I and Type II Scales Provide?

A recent paper by Luce (1972) poses the question, "What sort of measurement is psychophysical measurement?" One issue he raises is whether measurement of subjective variables like loudness, brightness, and heaviness is really analogous to fundamental measurement of physical quantities like mass, length, and duration. Basic to Luce's viewpoint is the observation that quantities measured physically interrelate with one another in physical laws. Typically these equations are multiplicative in form, in which each physical dimension appears raised to some simple power. An example is inverse-square laws. Do analogous formulations hold for psychophysical relations?

Both Type I and Type II psychophysical functions (and cross-modality matching functions) appear well approximated by power equations; considered in this way, those equations do resemble many physical

equations. But, as Luce points out, there has been no clear demonstration that the exponents of the psychophysical and sensory-physical power equations are wholly invariant from person to person. If they are not invariant, then psychophysical laws and cross-modality matching equations cannot be expressed as fixed relations among psychological and physical variables. Instead, the equations will have to be specific to the individual.

Perhaps it is premature to look seriously at the question that Luce raises. But it is valuable, I think, to make clear just what it is in psychophysics that is analogous to physics. The internal structure that physical laws provide to the system of physical dimensions is totally self-contained. The most direct analogue in the sensory realm is, therefore, neither the set of psychophysical equations (Type I or Type II) nor sensory-physical, cross-modality matching equations, but rather what I have termed psychosensory laws. The simple additivity of the loudnesses in tonal complexes and the additivity of loudnesses by the two ears provide two examples. These are laws that remain wholly within the psychological realm, just as physical laws remain wholly within the physical.

An even more interesting and theoretically more important example (from the scaling point of view) emerges from a study conducted by Stevens, Guirao, and Slawson (1965), who obtained magnitude estimates of three different attributes of sounds: subjective intensity (loudness), subjective compactness (density), and subjective spaciousness (volume). All three auditory dimensions had previously been shown to be (Type I) power functions of sound pressure; the size of the power-function exponent varies in a complex manner with sound frequency and does so in a different way for each dimension. But whereas these psychophysical relations are complicated, the psychosensory relation among the three attributes is simplicity itself: loudness equals the product of density and volume. The analogy to the physical interrelation among mass, density, and volume is clear. In the purely psychological domain, the units of subjective loudness, density, and volume are dimensionally consistent. Density, for instance, is unit loudness per unit volume. Thus, eventually it may well prove that Type I sensory measurement, at least, does provide a clear parallel to fundamental physical measurement.

REFERENCES

AIBA, T. S., & STEVENS, S. S. Relation of brightness to duration and luminance under light- and dark-adaptation. *Vision Research*, 1964, 4, 391-401.

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, 77, 153-170.

ANDERSON, N. H. Cross-task validation of functional measurement. *Perception & Psychophysics*, 1972, 12, 389-395.

ANGLIN, J. W., & MANSFIELD, R. J. W. On the brightness of short and long flashes. *Perception & Psychophysics*, 1968, 4, 161-162.

ATTNEAVE, F. Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 4. New York: McGraw-Hill, 1962. Pp. 619-659.

BANKS, W. P. A new psychophysical ratio scaling technique: Random production. *Bulletin of the Psychonomic Society*, 1973a, 1, 273-275.

BANKS, W. P. Reaction time as a measure of summation of warmth. *Perception & Psychophysics*, 1973b, 13, 321-327.

BANKS, W. P., & HILL, D. K. The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology*, in press.

BECK, J., & SHAW, W. A. The scaling of pitch by the method of magnitude estimation. *American Journal of Psychology*, 1961, 74, 242-251.

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

BOAKES, R. A. The bisection of a brightness interval by pigeons. *Journal of the Experimental Analysis of Behavior*, 1969, 12, 201-209.

BROCA, A., & SULZER, D. La sensation lumineuse en fonction du temps. *Comptes Rendus de l'Académie des Sciences*, 1902, 134, 831-834; 137, 944-946, 977-979, 1046-1049.

CHOCHOLLE, R. Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *L'Année Psychologique*, 1940, 41, 65-124.

CHURCHER, B. G., KING, A. J., & DAVIES, H. The measurement of noise with special reference to engineering noise problems. *Journal of the Institution of Electrical Engineers*, 1934, 75, 401-446.

CROSS, D. V. An application of mean value theory to psychophysical measurement. Unpublished manuscript, University of Michigan, 1965.

CROSS, D. V. Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, 1973, 14, 547-552.

CURTIS, D. W. Magnitude estimations and category judgments of brightness and brightness intervals: A two-stage interpretation. *Journal of Experimental Psychology*, 1970, 83, 201-208.

CURTIS, D. W., ATTNEAVE, F., & HARRINGTON, T. L. A test of a two-stage model for magnitude estimation. *Perception & Psychophysics*, 1968, 3, 25-31.

CURTIS, D. W., & RULE, S. J. Magnitude judgments of brightness and brightness difference as a function of background reflectance. *Journal of Experimental Psychology*, 1972, 95, 215-222.

DAWSON, W. E. Magnitude estimation of apparent sums and differences. *Perception & Psychophysics*, 1971, 9, 368-374.

EISLER, H. Subjective similarity in the continuum of heaviness with some theoretical and methodological considerations. *Scandinavian Journal of Psychology*, 1960, 1, 69-81.

EISLER, H. Empirical test of a model relating magnitude and category scales. *Scandinavian Journal of Psychology*, 1962, 3, 88-96.

EISLER, H., & EKMAN, G. A mechanism of subjective similarity. *Acta Psychologica*, 1959, 1, 1-10.

EKMAN, G., ENGEN, T., KÜNNAPAS, T., & LINDMAN, R. A quantitative principle of qualitative similarity. *Journal of Experimental Psychology*, 1964, 68, 530-534.

EKMAN, G., GOUDE, G., & WAERN, Y. Subjective similarity in two perceptual continua. *Journal of Experimental Psychology*, 1961, 61, 222-227.

FELTKELLER, R., ZWICKER, E., & PORT, E. Lautstärke, Verhältnislautheit, und Summenlautheit. *Frequenz*, 1959, 4, 108-117.

FLETCHER, H., & MUNSON, W. A. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, 1933, 5, 82-108.

FUORTES, M. G. F., & HODGKIN, A. L. Changes in times scale and sensitivity in the ommatidia of Limulus. *Journal of Physiology*, 1964, 172, 239-263.

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

GARNER, W. R. On the lambda loudness function, masking, and the loudness of multicomponent tones. *Journal of the Acoustical Society of America*, 1959, 31, 602-607.

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

GIBSON, R. H., & TOMKO, D. L. The relation between category and magnitude estimates of tactile intensity. *Perception & Psychophysics*, 1972, 12, 135-138.

GLASSER, L. G., McKENNEY, A. H., REILLY, A. H., & SCHNELLE, P. D. Cube-root color coordinate system. *Journal of the Optical Society of America*, 1958, 48, 736-740.

GRAHAM, C. H., & KEMP, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *Journal of General Psychology*, 1938, 21, 635-650.

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

HELM, C. E. Multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of America*, 1964, 54, 256-262.

HELLMAN, R. P., & ZWISLOCKI, J. J. Loudness function of a 1000-cps tone in the presence of a masking noise. *Journal of the Acoustical Society of America*, 1964, 36, 1618-1627.

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

HERRNSTEIN, R. J., & VAN SOMMERS, P. Method for sensory scaling with animals. *Science*, 1962, 135, 40-41.

INDOW, T., & KANAZAWA, K. Multidimensional mapping of Munsell colors varying in hue, chroma, and value. *Journal of Experimental Psychology*, 1960, 59, 330-336.

INDOW, T., & UCHIZONO, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *Journal of Experimental Psychology*, 1960, 59, 321-329.

JONES, F. N., & WOSKOW, M. H. Some effects of context on the slope in magnitude estimation. *Journal of Experimental Psychology*, 1966, 71, 177-180.

KATZ, M. S. Brief flash brightness. *Vision Research*, 1964, 4, 361-373.

KELLER, M. The relation between the critical duration and intensity in brightness discrimination. *Journal of Experimental Psychology*, 1941, 28, 407-418.

KRANTZ, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, 9, 168-199.

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

LEVELT, W. J. M., RIEMERSMA, J. B., & BUNT, A. A. Binaural additivity of loudness. *British Journal of Mathematical & Statistical Psychology*, 1972, 25, 51-68.

LEWIS, M. F. Category judgments as functions of flash luminance and duration. *Journal of the Optical Society of America*, 1965, 55, 1655-1660.

LUCE, R. D. What sort of measurement is psychophysical measurement? *American Psychologist*, 1972, 27, 96-106.

LUCE, R. D., & GREEN, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, 79, 14-57.

LUCE, R. D., & TUKEY, J. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, 1, 1-27.

MANSFIELD, R. J. W. Intensity relations in vision: Analysis and synthesis in a non-linear sensory system. Doctoral dissertation, Harvard University, 1970.

MANSFIELD, R. J. W. Brightness function: Effect of area and duration. *Journal of the Optical Society of America*, 1973, 63, 913-920.

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

MARKS, L. E. Visual brightness: Some applications of a model. *Vision Research*, 1972, 12, 1409-1421.

MARKS, L. E. *Sensory processes: The new psychophysics.* New York: Academic Press, 1974.

MARKS, L. E., & CAIN, W. S. Perception of intervals and ratios

for three prothetic continua. *Journal of Experimental Psychology*, 1972, 94, 6-17.

MARKS, L. E., & STEVENS, J. C. Individual brightness functions. *Perception & Psychophysics*, 1966, 1, 17-24.

MATIN, L. Critical duration, the differential luminance threshold, critical flicker frequency and visual adaptation: A theoretical treatment. *Journal of the Optical Society of America*, 1968, 58, 404-415.

McGILL, W. J. Loudness and reaction time: A guided tour of the listener's private world. *Acta Psychologica*, 1961, 19, 193-199.

MILLER, A. L., & SHELDON, R. Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology*, 1969, 81, 16-21.

MÜNSTERBERG, H. *Beiträge zur experimentellen Psychologie. III.* Freiburg: Mohr, 1890.

NACHMIAS, J., & STEINMAN, R. M. Brightness and discriminability of light flashes. *Vision Research*, 1965, 5, 545-557.

NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the O.S.A. Subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 1943, 33, 385-418.

PARDUCCI, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, 72, 407-418.

PARKER, S., & SCHNEIDER, B. Non-metric scaling of loudness and pitch using similarity and difference estimates. *Perception & Psychophysics*, 1974, in press.

PIÉRON, H. Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensities excitatrices. *L'Année Psychologique*, 1914, 20, 2-96.

PIÉRON, H. *The sensations: Their functions, processes and mechanisms.* New Haven: Yale University Press, 1952.

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

POLLACK, I. Neutralization of stimulus bias in the rating of grays. *Journal of Experimental Psychology*, 1965b, 69, 564-578.

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

RAAB, D. H. Magnitude estimation of the brightness of brief foveal stimuli. *Science*, 1962, 135, 42-44.

RAAB, D., FEHRER, E., & HERSHENSON, M. Visual reaction time and the Broca-Sulzer phenomenon. *Journal of Experimental Psychology*, 1961, 61, 193-199.

RAMSAY, J. O. Economical method of analyzing perceived color differences. *Journal of the Optical Society of America*, 1968, 58, 19-22.

REYNOLDS, G. S., & STEVENS, S. S. Binaural summation of loudness. *Journal of the Acoustical Society of America*, 1960, 32, 1337-1344.

RICHARDS, A. M. Monaural loudness functions under masking. *Journal of the Acoustical Society of America*, 1968, 44, 599-604.

RICHARDSON, L. F., & ROSS, J. S. Loudness and telephone current. *Journal of General Psychology*, 1930, 3, 288-306.

ROBINSON, D. W. The subjective loudness scale. *Acustica*, 1957, 7, 217-233.

ROWLEY, R. R., & STUDEBAKER, G. A. Monaural loudness-intensity relationships for a 1000-Hz tone. *Journal of the Acoustical Society of America*, 1969, 45, 1186-1192.

RSCHEVKIN, S. N., & RABINOVITCH, A. V. Sur la problème de l'estimation quantitative de la force d'un son. *Revue d'Acoustique*, 1936, 5, 183-200.

RULE, S. J., & CURTIS, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, 97, 305-309.

RULE, S. J., CURTIS, D. W., & MARKLEY, R. P. Input and output transformations from magnitude estimation. *Journal of Experimental Psychology*, 1970, 86, 343-349.

SAUNDERSON, J. L., & MILNER, B. I. A further study of ω space. *Journal of the Optical Society of America*, 1944, 34, 167-173.

SCHARF, B., & FISHKEN, D. Binaural summation of loudness: Reconsidered. *Journal of Experimental Psychology*, 1970, 86, 374-379.

SCHARF, B., & STEVENS, J. C. The form of the loudness function

near threshold. *Proceedings of the 3rd International Congress on Acoustics.* Amsterdam: Elsevier, 1959. Pp. 80-82.

SCHNEIDER, B. Multidimensional scaling of color difference in the pigeon. *Perception & Psychophysics,* 1972, **12,** 373-378.

SCHNEIDER, B., & LANE, H. Ratio scales, category scales, and variability in the production of loudness and softness. *Journal of the Acoustical Society of America,* 1963, **35,** 1953-1961.

SCHNEIDER, B., WRIGHT, A. A., EDELHEIT, W., HOCK, P., & HUMPHREY, C. Equal loudness contours derived from sensory magnitude judgments. *Journal of the Acoustical Society of America,* 1972, **51,** 1951-1959.

SHEPARD, R. Metric structures in ordinal data. *Journal of Mathematical Psychology,* 1966, **3,** 287-315.

SPERLING, G., & SONDHI, M. M. Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America,* 1968, **58,** 1133-1145.

STEVENS, J. C. Stimulus spacing and the judgment of loudness. *Journal of Experimental Psychology,* 1958, **56,** 246-250.

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

STEVENS, J. C., & HALL, J. W. Brightness and loudness as functions of stimulus duration. *Perception & Psychophysics,* 1966, **1,** 319-327.

STEVENS, J. C., & MACK, J. D. Scales of apparent force. *Journal of Experimental Psychology,* 1959, **58,** 405-413.

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

STEVENS, J. C., & MARKS, L. E. Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences,* 1965, **54,** 407-411.

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

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

STEVENS, S. S. The measurement of loudness. *Journal of the Acoustical Society of America,* 1955, **27,** 815-829.

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

STEVENS, S. S. On the psychophysical law. *Psychological Review,* 1957, **64,** 153-181.

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology,* 1959, **57,** 201-209.

STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961a. Pp. 1-33.

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

STEVENS, S. S. Duration, luminance, and the brightness exponent. *Perception & Psychophysics,* 1966, **1,** 96-100.

STEVENS, S. S. Issues in psychophysical measurement. *Psychological Review,* 1971, **78,** 426-450.

STEVENS, S. S. Perceived level of noise by Mark VII and decibels (E). *Journal of the Acoustical Society of America,* 1972, **51,** 575-601.

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

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

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

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

STEVENS, S. S., GUIRAO, M., & SLAWSON, A. W. Loudness, a product of volume times density. *Journal of Experimental Psychology,* 1965, **69,** 503-510.

STEVENS, S. S., & HARRIS, J. R. The scaling of subjective roughness and smoothness. *Journal of Experimental Psychology,* 1962, **64,** 489-494.

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

STEVENS, S. S., & STEVENS, J. C. The dynamics of visual brightness. Psychophysical Project Report PPR-246, Harvard University, 1960.

STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: A revised scale. *American Journal of Psychology,* 1940, **53,** 329-353.

STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America,* 1937, **8,** 185-190.

TEGHTSOONIAN, M. The judgment of size. *American Journal of Psychology,* 1965, **78,** 392-402.

TORGERSON, W. S. *Theory and method of scaling.* New York: Wiley, 1958.

VAUGHAN, H. G., JR., COSTA, L. D., & GILDEN, L. The functional relation of visual evoked response and reaction time to stimulus intensity. *Vision Research,* 1966, **6,** 645-656.

WARD, L. M. Category judgments of loudness in the absence of an experimenter-induced identification function: Sequential effects and power-function fit. *Journal of Experimental Psychology,* 1972, **94,** 179-184.

WARREN, R. M., & POULTON, E. C. Lightness of grays: Effects of background luminance. *Perception & Psychophysics,* 1966, **1,** 145-148.

WEISS, D. J. Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics,* 1972, **12,** 385-388.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology,* 1969, **82,** 52-63.

WRIGHT, H. Precision of color differences derived from a multidimensional scaling experiment. *Journal of the Optical Society of America,* 1965, **55,** 1650-1655.

YOUNG, F. W. Nonmetric scaling of line lengths using latencies, similarity, and same-different judgments. *Perception & Psychophysics,* 1970, **8,** 363-369.

NOTES

1. Krueger (1970) examined how subjects perceive combined lengths of sets of lines. Matches of single line lengths to apparent combined lengths suggested a nonlinear relation between subjective and objective length. It may be that the extensive continuum of length is, under some circumstances, conducive to a mode of perception nonlinearly related to Type I perception.

2. Perhaps, if the visual system does contain a two-stage feedback network like that outlined earlier, reaction time is determined at the first stage. Since the output from the first stage is predicted to be approximately proportional to the two-thirds power of luminance, there would again be a square-root relation between reaction speed and response magnitude.