5

Psychophysical Scaling[1,2]

R. Duncan Luce
University of Pennsylvania

Eugene Galanter
University of Washington

[1] *This work was supported in part by grant NSF G-17637 from the National Science Foundation to the University of Pennsylvania and by the Office of Naval Research, Contract NONR-551(37). Reproduction in whole or part is permitted for any purpose of the United States Government.*

[2] *We are indebted to Francis W. Irwin, Elizabeth F. Shipley, and S. S. Stevens for their helpful criticisms of a preliminary draft of this chapter. A special word of thanks is due Dr. Stevens, who, in addition to discovering several errors of fact and a number of stylistic blunders, forced us to reconsider some of our views on various aspects of scaling. His comments coupled with our own misgivings resulted in a complete revision of the section on magnitude estimation and the addition of some new theoretical work. Hopefully, this has produced some convergence of viewpoint.*

Contents

1. Similarity Scales 250
   1.1 The method of triads, 250
   1.2 A comparative judgment analysis: multidimensional scaling, 251
   1.3 A choice theory analysis, 253

2. Bisection Scales 256
   2.1 The method, 256
   2.2 A similarity analysis, 257
   2.3 A measurement analysis, 259

3. Category Scales 261
   3.1 The method, 261
   3.2 The mean category scale, 264
   3.3 Successive intervals and related scaling models, 269
   3.4 A choice analysis, 271

4. Magnitude Estimation Scales 273
   4.1 The method, 273
   4.2 The psychophysical law, 276
   4.3 The invariance of the scale, 278
   4.4 A probabilistic response theory, 283
   4.5 Form of the generalization function, 288
   4.6 Form of the psychophysical function, 291
   4.7 Relations to other experiments, 293

5. Distances 295
   5.1 A rationalization for distance, 296
   5.2 The embedding problem, 303

6. Conclusions 304

References 306

Psychophysical Scaling

In the preceding two chapters we have examined models for asymptotic choice behavior in a number of common psychophysical identification experiments. By the definition of an identification experiment, the perceptual problem was partly prejudged: certain physical orderings of the stimuli were assumed to correspond to the subject's perceptual orderings of them. For example, in loudness discrimination experiments the usual measure of physical intensity is assumed to order tones in the same way that the subject's perception of loudness does. The main effect of this assumption is to permit us to say whether or not a response to a stimulus presentation is correct, and so it seems acceptable to feed back information and to use payoffs.

The response theories so far proposed to describe behavior in such experiments all have two distinct classes of numerical parameters, one reflecting the effects of stimuli and the other, motivational biases. In testing these theories, it is necessary, among other things, to show that the stimulus parameters are in fact stimulus determined in the sense that they do not change when payoffs, presentation probabilities, and experimental designs are varied in certain ways. Once such a response theory is accepted, one must next determine just how the bias parameters depend upon the payoffs, the presentation probabilities, and whatever else they depend upon, and how the stimulus parameters depend upon physical properties of the stimuli. The latter relation is often called a *psychophysical scale.*

To some extent, we have already examined psychophysical scaling theories (Sec. 6, Chapter 3, and Sec. 2, Chapter 4), and in Sec. 2.1 of Chapter 4 we expressed some views on the general scaling problem which should be reread as background for this chapter. In addition to what we have already described, a number of other methods and models exist which attempt to treat the scaling of stimuli rather more directly and completely. These methods differ in two important respects from identification experiments. First, they can be used to organize a large part, if not all, of the sensible range of stimulation within a modality, not just some local region such as the neighborhood of the threshold or a two- or three-jnd interval about stimuli well above threshold. Second, the perceptual problem is no longer prejudged, and so neither payoff nor identification functions are involved. As a result, attention is directed almost exclusively to how subjects "organize" the stimuli according to some verbal instructions given by the experimenter and not to other features of the behavior. The philosophy underlying this approach is succinctly summarized in the following comments of S. S. Stevens.

> In a sense there is only one problem of psychophysics, namely, the definition of the stimulus. In this same sense there is only one problem in all of psychology —and it is the same problem. The definition of the stimulus is thus a bigger problem than it appears to be at first sight. The reason for equating psychology to the problem of defining stimuli can be restated thus: the complete definition of the stimulus to a given response involves the specification of all the transformations of the environment, both internal and external, that leave the response invariant. (1951, pp. 31–32.)

One consequence of not prejudging the perceptual problem is implicit in this quotation, namely a de-emphasis of the motivational factors which also influence behavior. Although Stevens mentions the "internal . . . environment," the fact of the matter is that people who do scaling experiments have not explicitly treated motivational questions. Yet, in the theories developed for identification experiments, stimuli and outcomes play complementary and equally important roles in determining the response. It is a little difficult to believe that the motivational factors have suddenly dropped from view just because we are certain that we do not understand the perceptual organization of the stimuli. Indeed, exactly the opposite seems more plausible. When the criterion for organizing the stimuli is uncertain to the experimenter, as for example when he asks a subject to make similarity judgments, it is probably equally vague to the subject, in which event his motives are likely to influence significantly his responses.

A closely related point is the fact that the experiments in question have to be somewhat modified before we can study the similarity perception of animals. We can ask a human subject which of two stimuli is more similar to a third or require him to group a set of stimuli into k equally spaced categories and usually he will comply without too much fuss, but with animals our only means of instruction is differential outcomes. For example, to study similarity judgments, we might first train the animal according to some more or less arbitrary identification function and then test their generalization to new stimuli during extinction trials (Herrnstein & van Sommers, 1962). Just how the results of such an experiment are related to those that we usually obtain from human beings in nonidentification experiments is an important research question about which little is known.

Given the data from a nonidentification, "perceptual" choice experiment, the usual procedure of analysis is this. One of the simpler response models for identification experiments, that is, one having no response bias parameters, is selected and is assumed to apply to a nonidentification design. The reasons for using the simpler models are that they are older and therefore better known, that they are easier to work with mathematically, and that an extra set of parameters for which there are no experimental counterparts can be a trifle embarrassing. On the assumption that the chosen response model is correct—often this can only be assumed because there are no experimental manipulations available with which to generate adequate tests—the stimulus parameters are calculated from the data. With these known for a number of stimuli from some extensive, homogeneous class of stimuli, the central question then is: what sort of "natural" organization do the scale values exhibit? If, for example, each stimulus is assigned a scale value, then we may ask, do these values, aside from sampling errors, stand in a simple functional relation to some physical measure of the stimuli? If scale values are assigned to pairs of stimuli, then we inquire whether the individual stimuli can be treated as points in some multidimensional space—Euclidean or otherwise—in such a way that distances in the space correspond approximately to the response-theory scale values.

This is what is done and what we shall describe in some detail in much of this chapter. What is not clear is why we have not yet evolved a somewhat more subtle approach using payoffs. For example, one might proceed in the following way. Let us assume that the effect of the vague instructions is to induce an unknown identification function in the subject and that part of our problem is to discover it. In general, any payoff function we use is going to be incompatible with it (see Sec. 3.3, Chapter 2, for a precise definition of compatibility), but different functions will be incompatible in different ways. These differences may give us some leverage on the problem. If we knew how incompatible payoffs and identification functions combine to generate responses, then the response data from a sufficient number of different payoff functions should permit us to "solve" for the unknown identification function. Just how many different experiments are needed to get a determinate solution depends, of course, upon the exact mathematical nature of the response theory, that is, upon exactly how the subject compromises his perceptions and his motivations.

The only difficulty in carrying out this program is that we do not know what response theory to use when the identification function and the payoffs are incompatible. Having noted this, however, it is clear what to do: we must perform identification experiments with incompatible payoffs, the goal being to work out suitable response theories that parallel those we now have for compatible situations. In all likelihood these theories will generalize the ones for compatible payoffs. Once such a theory is developed and tested, we can assume it applies when the identification function is unknown, solve for this unknown function, and then test the adequacy of the theory in the new context by predicting behavior for the other payoff functions.

Although this approach seems sensible, no work along these lines appears to have been reported. In the existing scaling studies identification functions are not defined, and neither payoffs nor information feedback are employed. Several quite different methods are of current interest and are discussed in the remainder of the chapter. We present them in what, it seems to us, is an order of decreasing familiarity. No other more compelling organization is apparent. At first we deal with experiments and theories that closely parallel those discussed in the preceding chapters, and then we move on to others that are more novel and less well understood.

1. SIMILARITY SCALES

1.1 The Method of Triads

Certain identification experiments are thought to yield information about the subjective similarity of pairs of stimuli, even though no direct judgments of similarity are made. For example, the choice theory analysis of complete identification experiments (Sec. 1.2, Chapter 3) led to scale values $\eta(s, s')$, which were interpreted as a possible measure of the similarity between pairs of stimuli. In addition to these theoretical interpretations, one can ask the subjects to make explicit similarity judgments. Because we have no precise, nonarbitrary notion about what psychological similarity might mean in terms of physical properties of the stimuli, we are forced to use nonidentification experimental designs. In this section we discuss in detail the one known as the method of triads; save for the absence of an identification function, it resembles the forced-choice discrimination design.

Let a, x, $y \in \mathscr{S}$. A typical stimulus presentation is $\langle x, y, a\rangle$, and the subject is instructed to report which of the first two stimuli in the presentation seems to him "more similar" to the third, the so-called *reference stimulus*. (Of course, the reference stimulus can be located in any of the three positions, and where it is may very well alter the experimental findings to some degree. For our purposes it will be convenient to locate it in the last position, realizing that this is merely a notational convenience.)

In general, then,

$$S \subseteq \mathscr{S}^3 \quad \text{and} \quad R = \{1, 2\}.$$

If we confine our attention to those experiments, or to those parts of one, in which there is a single reference stimulus a, then

$$S \subseteq \mathscr{S}^2 \times \{a\}.$$

Now the close parallel to simple discrimination experiments is obvious; the only differences are that the reference stimulus is added to each presentation and, of course, that the subject is asked to make a judgment of similarity, not relative magnitude.

Obviously, the triad design is readily generalized to one of choosing which of k stimuli is most similar to a, in which case

$$S \subseteq \mathscr{S}^k \times \{a\} \quad \text{and} \quad R = \{1, 2, \ldots, k\}.$$

The word "similar" used in the instructions is vague, and it is left that way because neither the experimenter nor the subject can verbalize very precisely what he means by it. Nonetheless, subjects respond nonrandomly when instructed in this way. That reproducible data can arise from a vague criterion should not surprise us when we think of how often we use equally vague criteria in everyday life, but in the long run a science is not likely to let reproducibility alone substitute for well analyzed and controlled experimental designs.

The responses in the triad design are assumed to be generated from a probabilistic process having the basic conditional probabilities

$$p(i \mid \langle x, y, a\rangle), \qquad (i = 1, 2),$$

where

$$p(1 \mid \langle x, y, a\rangle) + p(2 \mid \langle x, y, a\rangle) = 1.$$

If the order of presentation does not matter, then as in discrimination work we can write

$$p(x, y; a) = p(1 \mid \langle x, y, a\rangle) = p(2 \mid \langle y, x, a\rangle).$$

The generalization to more stimuli is clear.

A somewhat more general procedure used to study similarity is the method of tetrads in which

$$S \subseteq \mathscr{S}^4 \quad \text{and} \quad R = \{1, 2\},$$

and the subject is asked to judge whether the first or second pair of stimuli presented is more similar. Suppes and Zinnes discuss models for this experiment in Secs. 3.3, 3.4, and 4.4, of Chapter 1; we shall not go into them here.

1.2 A Comparative Judgment Analysis: Multidimensional Scaling

Assume for the moment that, as in previous Thurstonian models we have examined, there is a random variable $\mathbf{X}$ in the real numbers which represents the effect of stimulus x and that the random variables associated with different stimuli assume values on the same numerical scale. If the order of presentation does not matter, the obvious decision rule for the method of triads is the following:

Stimulus x rather than y is judged to be more similar to the reference stimulus a if and only if $|\mathbf{X} - \mathbf{A}| < |\mathbf{Y} - \mathbf{A}|$, *where the vertical bars denote the absolute value of the number between them.*

Thus

$$p(x, y; a) = \Pr(|\mathbf{X} - \mathbf{A}| < |\mathbf{Y} - \mathbf{A})|.$$

Note that the quantity $|\mathbf{X} - \mathbf{A}|$ can be interpreted as a distance random variate, $\mathbf{D}(x, a)$, that represents the momentary psychological distance between x and a on the decision continuum. This immediately suggests a multidimensional generalization of the Thurstone model, in which the random variables assume values in a k-dimensional Euclidean vector space of effects. As early as 1938 M. W. Richardson suggested that such models would be necessary to provide adequate representations of complex stimulus domains.

Let $\mathbf{X}$ denote a random vector assuming values in a k-dimensional Euclidean vector space and let its components be $\mathbf{X}_i$, $i = 1, 2, \ldots, k$. If the usual Euclidean distance measure is assumed,

$$\mathbf{D}(x, a)^2 = \sum_{i=1}^{k} (\mathbf{X}_i - \mathbf{A}_i)^2,$$

then the decision rule becomes

Stimulus x rather than y is judged to be more similar to the reference stimulus a if and only if $\mathbf{D}(x, a) < \mathbf{D}(y, a)$.

So

$$p(x, y; a) = \Pr[\mathbf{D}(x, a) < \mathbf{D}(y, a)].$$

As pointed out by Suppes and Zinnes in Sec. 4.4 of Chapter 1, the square of the distance $\mathbf{D}(x, a)$ has a noncentral χ^2 distribution, provided that the components $\mathbf{X}_i$ and $\mathbf{A}_i$ have normal distributions with the same variance. This fact, which makes matters rather more complicated than in previous Thurstonian models that we have examined, seems to have been overlooked in the published literature.

Torgerson (1952, 1958) attempted to bypass this complication by stating directly an analogue of the equation of comparative judgment, namely,

$$d(x, a) - d(y, a) = Z(x, y; a), \qquad (1)$$

where $Z(x, y; a)$ is the unit normal deviate corresponding to $p(x, y; a)$.

Implicitly, this postulates that the difference $\mathbf{D}(x, a) - \mathbf{D}(y, a)$ is a normally distributed random variable with mean $d(x, a) - d(y, a)$ and unit variance. Just where Eq. 1 comes from, aside from being the purely formal analogue of the discrimination model, is not clear. As Suppes and Zinnes point out, it certainly does not make sense to assume that $\mathbf{D}(x, a)$ is normally distributed, because it must have the distance property $\mathbf{D}(x, a) \geqslant 0$. Thus Torgerson's multidimensional scaling model is *ad hoc* in the sense that it does not derive from the same basic considerations as the other Thurstonian models.

Given that Eq. 1 holds, Torgerson (1952, 1958) presented a least squares solution to the problem of estimating the values $d(x, a) - d(y, a)$. Because only differences of mean distances are estimated, the individual means are determined up to a positive linear transformation. This creates what is known as the problem of the additive constant. Because distances must form a ratio scale, the additive constant of the linear transformation is not in fact a free parameter, but rather it must have a fixed value, which, for some purposes, we must estimate.

Messick and Abelson (1956) proposed a general iterative solution to the problem, which is based, in part, upon an embedding theorem of Young and Householder (1938) (see Sec. 5.2). The details of the method are described in Torgerson (1958). For the unidimensional case, Torgerson (1952) gave a simple least squares solution which rests upon the following observation. Because the true distances must satisfy

$$d(x, z) = d(x, y) + d(y, z),$$

the calculated distances

$$d'(x, y) = d(x, y) + c,$$

which differ from the true ones by the additive constant c, must satisfy

$$\begin{aligned} d'(x, y) + d'(y, z) - d'(x, z) &= d(x, y) + c + d(y, z) + c - d(x, z) - c \\ &= c. \end{aligned}$$

Thus, if the data were error free, c would be determined. When they are not error free, Torgerson's procedure gives a "best" estimate of c in the least squares sense.

1.3 A Choice Theory Analysis

Because of the formal parallel between discrimination and similarity designs (Sec. 1.1), a choice theory analysis follows almost immediately if we reinterpret the unbiased discrimination model described in Sec. 3.2 of Chapter 4. Specifically, if a denotes the reference stimulus, T, the set of comparison stimuli, and $x \in T$, then the basic response probabilities of the similarity experiment are of the form $p_T(x, a)$. Thus, with a held fixed, the choice axiom may be written as before, and so a ratio scale v exists for each a. The typical scale value can be written in the form $v(x, a)$. This dependence upon two stimuli makes these parameters formally similar to those that arose in the choice analysis of complete identification experiments (Sec. 1.2, Chapter 3), and Luce (1961) proposed that the same assumptions be investigated:

Assumption 1. *For all* $x, y \in \mathscr{S}$, $v(x, y) = v(y, x)$.

Assumption 2. *For all* $x \in \mathscr{S}$, $v(x, x) = 1$.

Assumption 3. *For all* $x, y, z \in \mathscr{S}$, $v(x, z) \geqslant v(x, y)\, v(y, z)$.

Put another way, he suggested assuming that $-\log v$ is a distance measure.

The first of these assumptions, that concerning symmetry, is most important because it says that there is a single scale, not a collection of unrelated ones. If we let $p(x, y; z)$ denote $p_{\{x,y\}}(x, z)$, it is easy to show that Assumption 1 is equivalent to the (in principle) testable statement

$$p(x, y; z)\, p(y, z; x)\, p(z, x; y) = p(x, z; y)\, p(z, y; x)\, p(y, x; z).$$

The situation most carefully examined by Luce involves a strengthening of Assumption 3 so that $-\log v$ acts like distance on a line; presumably this restricted model can, at best, apply to physically unidimensional continua. Let us say that stimulus y is *between* stimuli x and z if, when z is the reference stimulus, z is more often judged similar to y than to x and, when x is the reference stimulus, x is judged more similar to y than to z, that is,

$$p(y, x; z) > \tfrac{1}{2} \quad \text{and} \quad p(y, z; x) > \tfrac{1}{2}.$$

Assumption 3′. *For all* $x, y, z \in \mathscr{S}$ *such that* y *is between* x *and* z, *then* $v(x, z) = v(x, y)\, v(y, z)$.

The main conclusion that has been derived from Assumptions 1, 2, and 3′, concerns the plot of $p(a, b; x)$ as a function of x, assuming that the stimuli differ only along one physical dimension. Note the reversal of viewpoint that has occurred. We began by thinking of the reference stimulus as a fixed quantity and the comparison stimuli as experimental variables; now we propose to think of the comparison stimuli as fixed and the reference as the variable. Suppose on the physical continuum that $a < b$. The result says that for $x \leqslant a$, $p(a, b; x)$ has a constant value, say K, and that for $x \geqslant b$ it has the constant value $1 - K$. For $a \leqslant x \leqslant b$, there is some (presumably continuous) transition from K to $1 - K$ (see Fig. 1). This transition function does not depend upon a and b independently but rather is associated with what may be called their midpoint. Specifically, we say that stimulus $\overline{ab}$ is the *midpoint* of a and b if $p(a, b; \overline{ab}) = \tfrac{1}{2}$. Now, if c

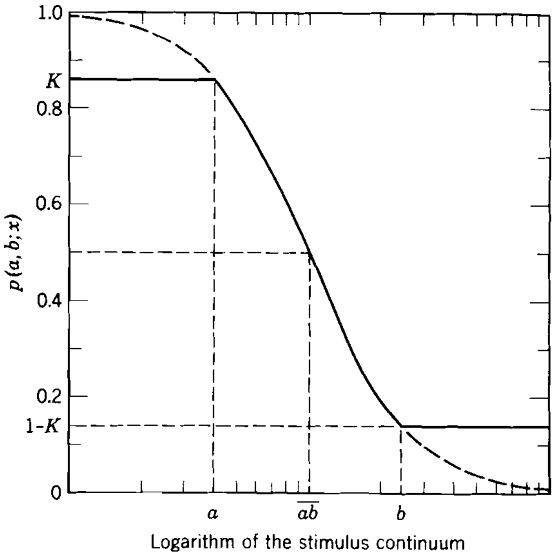

Fig. 1. The transition function derived from a choice theory analysis of similarity judgments. Adapted by permission from Luce (1961, p. 158).

and d, $c < d$, are two different stimuli such that $\overline{cd} = \overline{ab}$, it can be shown that when a, $c < x < b$, d, then $p(a, b; x) = p(c, d; x)$. That is to say, if two pairs of comparison stimuli have the same midpoint, then the two functions coincide in the region of overlap defined by the pairs of stimuli.

No empirical research has yet been performed to test this model. Whether or not it is correct, it will be interesting to develop empirical plots of $p(a, b; x)$ versus x simply to see what they are like.

It is of interest also to inquire about the relation between the stimulus-scale values obtained from the analysis of recognition data (Chapter 3) and those obtained from the analysis of similarity data gathered under the same experimental conditions and their relation to the scale values calculated from discrimination data. From a rather questionable assumption, Luce (1961) showed that

$$v(x, y) = \begin{cases} \dfrac{v(x)}{v(y)} & \text{if} \quad v(x) \leqslant v(y) \\ \dfrac{v(y)}{v(x)} & \text{if} \quad v(x) \geqslant v(y), \end{cases}$$

where the two-place v denotes the similarity scale value and the one-place v, the discrimination value. Formally, this same assumption was invoked in Sec. 7.3 of Chapter 3 in an attempt to account for some of the information theory results. The only difference is that the two-place scale value there denoted the recognition experiment scale value $\eta(x, y)$. We suspect that the weaker assumption

$$v(x, y) = \begin{cases} \left[\dfrac{v(x)}{v(y)}\right]^{\beta} & \text{if} \quad v(x) \leqslant v(y) \\ \left[\dfrac{v(y)}{v(x)}\right]^{\beta} & \text{if} \quad v(x) \geqslant v(y), \end{cases}$$

in which β is a parameter to be estimated from the data, is far more likely to receive support. Note that for x, y, and z, with $v(x) < v(y) < v(z)$, then either assumption implies:

$$v(x, z) = v(x, y)\, v(y, z).$$

2. BISECTION SCALES

2.1 The Method

The bisection design is uniquely different from anything else in psychophysics that we have discussed. As it is usually performed, the response literally involves the selection of a stimulus. Consider a stimulus set in which the stimuli differ on one physical dimension, such as sound intensity. An ordered pair of stimuli is presented to the subject, who adjusts the gain of a third presentation until, in his opinion, this variable tone has a subjective loudness that "bisects" the loudnesses of the fixed pair of tones. In practice, there are various ways to make this adjustment. In one of the most common the subject first chooses a gain setting and then listens to the ordered triplet $\langle a, x, b\rangle$, in which a and b are the fixed tones and x is the one he selected. Having heard the triplet, the subject decides whether he likes his setting; if he does not, he resets x, listens, and so on, until he is satisfied that his response stimulus "bisects" a and b. Observe that this selection of a stimulus is utterly different from that in any experiment previously described; in some that we have studied the responses identified one of the presented stimuli as larger, more similar, etc., than the others, but the subjects's choice was restricted to one of the stimuli presented. The whole stimulus set was not available.

In another method of studying bisection the experimenter selects the triples and asks the subject whether the test stimulus is above or below his bisection point. The 50 per cent point on the resulting psychometric function is taken to be the bisection stimulus. It is not clear that the two methods will yield the same results, but on the assumption that they do, then bisection can be interpreted as a special case of a similarity judgment. An analysis based upon this assumption is given in Sec. 2.2.

It should be noted in passing that bisection can also be considered as a special case of what S. S. Stevens (1958a) has called category production, which is the logical counterpart of category estimation discussed in Sec. 3. Because we know of no theory for the general case, we shall confine our attention to bisection.

If $\mathscr{S}$ denotes the set (usually a continuum) of stimuli, physically ordered by the relation $\geqslant$, then $S \subseteq \mathscr{S}^2$ and $R = \mathscr{S}$ in this design. The basic response data are presumed to be generated by conditional probabilities (or densities, as the case may be) of the form $p(x \mid \langle a, b \rangle)$, where $x \in R = \mathscr{S}$ and $\langle a, b \rangle \in S$. If $\mathscr{S}$ is made discrete by the design of the equipment, then in principle it is feasible to estimate these probabilities; if $\mathscr{S}$ is continuous, then parameters of the density function can be estimated.

As in much psychophysical work, the order of presentation matters. For intensive (prothetic) continua, the mean bisection value $\bar{x}$ in the ascending series $\langle a, x, b \rangle$, is consistently and appreciably different from the mean $\bar{y}$ in the descending series $\langle b, y, a \rangle$. Because of a superficial analogy to a well-known physical phenomenon, this response bias has been called *hysteresis*. Examples of it are shown in S. S. Stevens (1957).

No truly probabilistic model has yet been proposed for bisection data, the main reason being that three stimuli are involved—two in the presentation and one in the response. For the choice model, this leads to scale values of the form $v(x; a, b)$, and so some drastic simplifying assumptions are needed. For the discriminal dispersion model, one has to deal with the three random variables, **A**, **X**, and **B**. Presumably the decision rule would be something of the form: there exist positive constants c and d such that whenever

$$\frac{\mathbf{A} + \mathbf{B}}{2} - c \leqslant \mathbf{X} \leqslant \frac{\mathbf{A} + \mathbf{B}}{2} + d$$

the subject accepts x as the bisection value. If one were willing to postulate how the subject would alter his setting of the response stimulus as a function of the previously observed **A**, **X**, and **B**, then it would be possible to calculate the distribution of adjustments until the process terminated. Because the number of adjustments made is just as observable as the actual choice, such a model could be tested in some detail.

The only models we discuss here are essentially deterministic in nature.

2.2 A Similarity Analysis

If we assume that the subject interprets "bisect" to mean "equally similar to," in the sense of the method of triads, then any model for that method also is a model for bisection. What is assumed is that the subject adjusts the reference stimulus until he is satisfied that it is equally similar to both a and b. Because of the probabilistic nature of the similarity models, his choice of a bisection point must vary from trial to trial, but the mean value $\bar{x}$ might be defined by the property

$$p(a, b; \bar{x}) = \tfrac{1}{2}.$$

Assuming a similarity model with no response biases, which we know cannot be precisely correct because of the hysteresis effect, the bisection point then coincides with what we called the midpoint $\overline{ab}$ in Sec. 1.3. If in the choice model we suppose that there exists a constant β such that

$$v(x, y) = \begin{cases} \left(\dfrac{x}{y}\right)^{\beta} & \text{if} \quad x \leqslant y \\ \left(\dfrac{y}{x}\right)^{\alpha} & \text{if} \quad x \geqslant y, \end{cases}$$

then from

$$\begin{aligned} p(a, b; \overline{ab}) &= \tfrac{1}{2} \\ &= \frac{1}{1 + v(b, \overline{ab})/v(a, \overline{ab})}, \end{aligned}$$

and from the assumption that $a < \overline{ab} < b$ it follows that

$$\begin{aligned} \left(\frac{a}{\overline{ab}}\right)^{\beta} &= v(a, \overline{ab}) \\ &= v(b, \overline{ab}) \\ &= \left(\frac{\overline{ab}}{b}\right)^{\beta}. \end{aligned}$$

If $b < \overline{ab} < a$, then similarly

$$\left(\frac{\overline{ab}}{a}\right)^{\beta} = \left(\frac{b}{\overline{ab}}\right)^{\beta}.$$

Thus, in either case, $\overline{ab} = (ab)^{1/2}$, that is, the midpoint is predicted to be the geometric mean of the stimulus values that are bisected. In logarithmic—for example, decibel—measures the midpoint is predicted to be the arithmetic mean of the given stimulus values. This is empirically incorrect (S. S. Stevens, 1957), both because of the hysteresis effect and because both midpoints are above the geometric mean.

In Luce (1961) the similarity model of Sec. 1.3 is generalized to one having response biases, which overcomes the difficulties just described.

Although it seems plausible that bisection is a special case of a similarity judgment, experiments are definitely needed to test this hypothesis.

2.3 A Measurement Analysis

Pfanzagl (1959a,b), extending and reinterpreting Aczél's (1948) axiomatization of mean values, has created an interesting measurement axiom system, specializations of which yield a number of familiar measurement models for different subject matters. For example, it includes the classic models for the measurement of mass and of length, the von Neumann-Morgenstern axioms for utility, as well as a possible model for the measurement of sensation based upon bisection. We first present the bisection specialization of Pfanzagl's axioms and the resulting representation and uniqueness theorem; then we discuss the interpretation. Let $\mathscr{S}$ denote the set of stimuli, physically weakly ordered by $\geqslant$.

Axiom 1 (Existence). *For every* $x, y \in \mathscr{S}$ *there exists a unique element* $B(x, y) \in \mathscr{S}$*, which is interpreted as the bisection point of* x *and* y.

Axiom 2 (Monotonicity). *If* $x \lessgtr x'$*, then, for all* $y \in \mathscr{S}$, $B(x, y) \lessgtr B(x', y)$.

Axiom 3 (Continuity). *B is a continuous function in both of its arguments, which is to say that* $\{x \in \mathscr{S} \mid B(x, b) > a\}$, $\{x \in \mathscr{S} \mid B(x, b) < a\}$, $\{x \in \mathscr{S} \mid B(b, x) > a\}$, *and* $\{x \in \mathscr{S} \mid B(b, x) < a\}$ *are all topologically open sets for every* $a,b \in \mathscr{S}$.

Axiom 4 (Bisymmetry). *For all* $w, x, y, z \in \mathscr{S}$, $B[B(w, x), B(y, z)] = B[B(w, y), B(x, z)]$.

Axiom 5 (Reflexivity). *For all* $x \in \mathscr{S}$, $B(x, x) = x$.

Axioms 1 and 3 are largely technical in nature and need not be discussed. Axioms 2 and 5 are most plausible, and it seems unlikely that empirically they will be shown to be false. This leaves in doubt only Axiom 4, which contains most of the mathematical power of the system. Graphically, the various quantities involved in Axiom 4 are shown in Fig. 2. The assertion is that the two bisections of bisection points are the same. No thorough experimental investigation of this axiom has ever been made, but Pfanzagl (1959b) refers to studies of special cases for pitch, in which it seems to be sustained, and for loudness, in which it may not be.

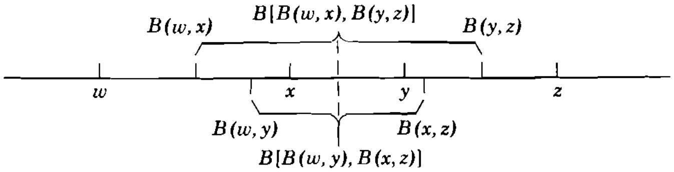

Fig. 2. A graphical representation of the quantities involved in Pfanzagl's Axiom 4.

Theorem 1. *If Axioms* 1 *to* 5 *hold, then there exists a real valued function u on* $\mathscr{S}$ *and a real number* δ, $0 < \delta < 1$, *such that*

1. *u is a continuous function;*
2. *u is a strictly monotonic function, that is, if $x < y$, $u(x) < u(y)$;*
3. *$u[B(x, y)] = \delta u(x) + (1 - \delta)u(y)$;*
4. *u is unique up to a positive linear transformation, that is, it is an interval scale.*

If, in addition, B is symmetric (commutative) in the sense that $B(x, y) = B(y, x)$, then $\delta = \frac{1}{2}$.

We shall not attempt to prove this result here; see Pfanzagl (1959a) for a full proof and (1959b) for a less complete one.

Because interval scales of "sensation" previously have appeared to be logarithmic functions of physical intensity, it is reasonable again to investigate the assumption that

$$u(x) = \alpha \log x + \beta,$$

where x is now both the physical magnitude and the name of the stimulus. It is easy, then, to show that

$$B(x, y) = x^{\delta}y^{1-\delta},$$

and so

$$B(x, y) = B(y, x)\left(\frac{x}{y}\right)^{2\delta-1}$$

Thus, for $\delta \neq \frac{1}{2}$, this model permits a hysteresis effect. For $\delta = \frac{1}{2}$, the bisection point is again the geometric mean of the given stimulus value.

This treatment of bisection has two major drawbacks. First, the data strongly suggest that a probabilistic, not a deterministic, model is needed. Of course, one can treat the deterministic analysis as an approximation to the probabilistic, for example, by letting

$$B(x, y) = \int_0^\infty zp(z \mid \langle x, y \rangle)\, dz,$$

but then the full probability process remains unanalyzed. Second, it is questionable whether the behavior is sufficiently invariant under various experimental manipulations to treat the phenomenon as a form of fundamental measurement, as this axiom system does. There is little doubt that the behavior can be altered by means of payoffs, and it is far from evident that the axioms will hold up under such changes. The most critical one, of course, is Axiom 4, and we suspect that it will not fare well when strong experimentally induced response biases exist. Despite the fact that many psychophysicists believe that they are in the business of discovering fundamental measures of sensation, response models, yielding derived measures, rather than fundamental measurement models, seem much more appropriate for psychophysical phenomena. The reason simply is that factors intrinsic to the experiment other than the stimuli importantly influence the responses. It is as though one tried to measure current without being aware of factors such as the area and temperature of a conductor, both of which affect its resistance and so the current flow. A good theory stating the relations among the relevant factors makes the accurate measurement of any one feasible; without such a theory, one only can try to hold the other variables constant, which may not be easy to do.

3. CATEGORY SCALES

3.1 The Method

In much the same sense that similarity experiments are analogous to discrimination designs, category experiments have some formal resemblance to those of recognition. Most of the recognition experiments that we discussed in Chapter 3 were complete identification designs, but it is clear what one would mean by a partial recognition design: $S = \mathscr{S}$ and certain responses are correct identifications for more than one stimulus presentation. The category experiments are analogous to this, except that no identification function, partial or otherwise, is specified by the experimenter.

Category methods are generally employed only when the stimuli can reasonably be considered to be ordered; for example, when they differ on only one physical dimension. Because the stimuli are ordered, it is reasonable to use responses that are also ordered in some way. Often the first m integers are used for the responses and the ordering is the natural one. Other response labels, such as the first m letters of the alphabet or descriptive adjectives, are sometimes employed. The subject is instructed to assign the "smallest" (weakest, lightest, darkest, etc.) stimulus to the first category, the "largest" to the mth category, and to use the other response categories so that his subjective impressions of the distances between successive categories is the same, that is, so that the categories are *equally spaced subjectively*. Because these instructions are vague, just as the similarity and bisection ones are, no identification function is assumed to be known.

It is generally felt that we are not demanding much more of the subject when we ask for category judgments rather than for similarity ones. If we believe that he can tell whether a is more similar to x than it is to y, then it should be possible for him to group stimuli into classes of comparable similarity. One suspects, however, that the meaningfulness of the obtained data depends upon the degree to which the subject understands what it is that he is being asked to do. Therefore, in testing the feasibility, reliability, and coherence of methods of this type, experimenters have generally first worked with simple physical dimensions, such as sound intensity, which they feel are relatively well understood both by the subjects and themselves. Later, the methods were extended to stimuli that have no known ordering other than a subjective one. Examples are the degrees of favorableness toward the church which are exhibited by certain statements and the degrees of intelligibility which are exhibited by handwriting samples. Formally, the nature of the stimuli makes no difference: once the relative frequencies are obtained, the models proceed without reference to the meaningfulness of the data. Substantively, there may well be important differences. For our purposes, it suffices and simplifies matters to consider only relatively simple physical stimuli.

The initial exploration of category methods was undertaken by experimenters primarily interested in discrimination (Titchener, 1905, Wever & Zener, 1928). They introduced, as a modification of the method of constant stimuli, what is called the *method of single stimuli.* It amounts to omitting the standard stimulus, so that only a single stimulus is presented on each trial. The subject's task is to judge whether a presentation is "loud" or "soft" or, in a variant, whether it is "loud," "medium," or "soft." After a few trials, during which the subject becomes acquainted with the range of stimuli involved, his responses settle down to "asymptotic" levels. It was found that psychometric functions generated in this way are quite similar to those generated by the method of constant stimuli (Fernberger, 1931). It is almost as if a subject defined his own standard stimulus for the given set of comparisons and that he was able to hold this image reasonably well fixed during the course of the experiment.

No identification functions were assumed in these studies, hence no information feedback or payoffs were used. Payoffs could have been used had the experimenter selected an arbitrary point on the continuum to separate loud from soft, but at the time this was considered inappropriate. Today, it is not so clear that payoffs should not be used. To be sure, the data for just one arbitrary cutpoint would not hold much interest, but those for several cutpoints from subjects judging the same set of stimuli could very well reveal what compromise the subject is making between his perceptions and the arbitrary feedback.

The method of single stimuli, although initially introduced only as a more rapid version of the method of constant stimuli, has certain important features of its own. It is easily adapted to yield nontrivial information over large ranges of stimulation by increasing the number of response categories. As early as 1898, E. P. Sanford (see Titchener, 1905, p. 82) had

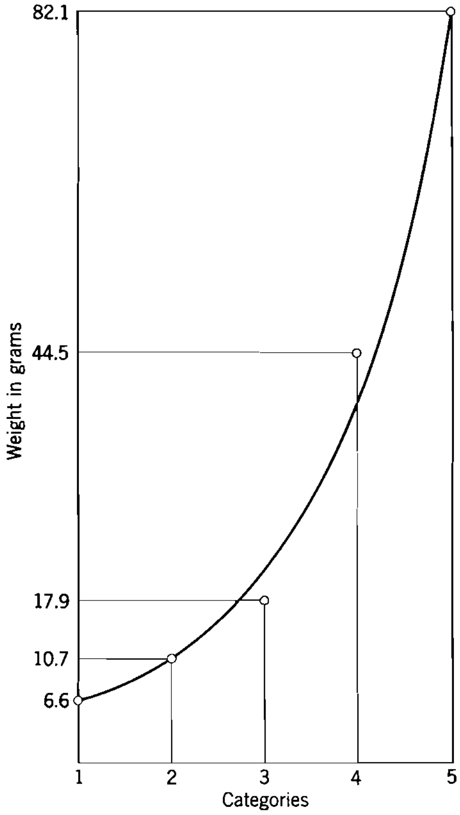

Fig. 3. The results of Sanford's weight-lifting experiment. Observe that the category judgments are plotted as the abscissa and the stimulus values are plotted as the ordinate. Adapted by permission from Titchener (1905).

experimental psychology students sort envelopes containing different weights into five categories of increasing weights. Category 1 was to be used for those that were lightest and 5 for those that were heaviest. The resulting plot of average stimulus weight against category number, shown in Fig. 3, was interpreted as a demonstration of Fechner's law. The observed curve is so close to Fechner's logarithmic law that Titchener claimed that the students had defined the categories so that they contained equal numbers of jnds. This idea for the definition of the categories was later adapted to serve as the basis of a Thurstonian theory of category judgments (Sec. 3.3).

Today, the following general procedure is used. The experimenter selects a set of m stimuli—usually m is about 20 but sometimes it is as large as 100—which he presents in random order to the subject. (Sometimes they are presented simultaneously, when this is feasible, but we shall confine our attention to trial-by-trial presentations of single stimuli.) To each presentation the subject responds by choosing one of k categories. Usually k is an odd number in the range from 5 to 11. When the stimuli differ in only one physical dimension, the instructional problem is relatively simple. The smallest stimulus is presented, and the subject is told that it is the smallest and is therefore a member of category 1; the largest is presented, and he is told that it is the largest and is therefore a member of category k; finally, he is told that he is to use the remaining categories as if they were equally spaced along the sensation continuum between these two extreme stimuli. Just what this means to subjects is not clear; there is some indication that they may interpret it to mean that the categories should be used equally often. For example, the assignments to categories are far from invariant when everything else is held fixed and the presentation probabilities are varied.

The data are the relative frequencies that response category r is used when stimulus presentation s is presented, and these frequencies are treated as estimates of underlying conditional probabilities $p(r \mid s)$. When the response categories are the first k integers and the stimuli are ordered, we usually denote a typical response by j and a typical stimulus by i and write $p(j \mid i)$.

3.2 The Mean Category Scale

The simplest analysis of the data involves calculating the mean category assignment for each stimulus and calling this number a "sensation scale value." At a theoretical level the scale is $u(s) = \sum_{j=1}^{k} jp(j \mid s)$. When the stimuli are presented many times to individual subjects, the mean can be calculated over presentations for each subject separately, as shown in Fig. 4 for an $m = 14$, $k = 7$ design, using white noise stimuli separated by five decibel steps of intensity. When each stimulus is presented just once to each subject, the mean is calculated over subjects. By and large, the data for individual subjects differ so little that means calculated over groups of subjects are considered adequate (see below, however, for an objection to this procedure).

The variety of stimulus domains that can be quickly explored and the ease with which various experimental manipulations can be evaluated by category methods has made this analysis into mean category judgments very popular. To the theorist, however, the whole business is a bit

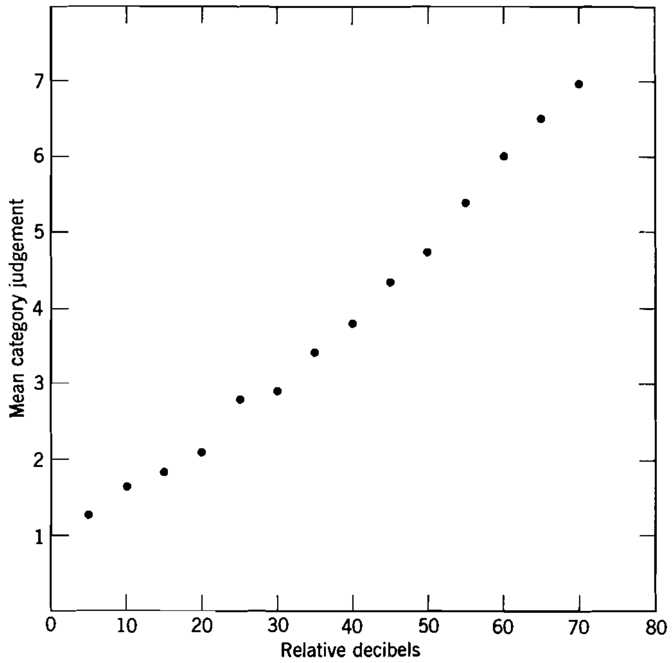

Fig. 4. The mean category judgments for a single subject for 20 independent irregular presentations of 14 white noise stimuli. The abscissa is labeled in relative decibels. Unpublished data of Eugene Galanter.

hair-raising. To calculate the means of category *labels*, to plot them against physical measures of the stimuli, and then to discuss the form of the resulting function strikes him as close to meaningless. Because there is nothing about the procedure to prevent one from labeling the categories by any other increasing sequence of numbers, we can by our choice of labels produce any arbitrary monotonic function of the physical stimuli we choose. What then can a particular one of these scales mean?

Although we do not think that the absolute form of the obtained function using the first k integers as labels has any meaning, the occurrence or nonoccurrence of changes in that function when various experimental parameters are changed may be a convenient way to summarize this class of empirical results.

If we use different ordered sets to label the responses, for example, names like soft, medium, and loud, letters of the alphabet, or different buttons in a line to depress, then we can make the natural identifications with the first k integers to calculate scale values. If everything save the labeling is held constant, then in general the data suggest that the function

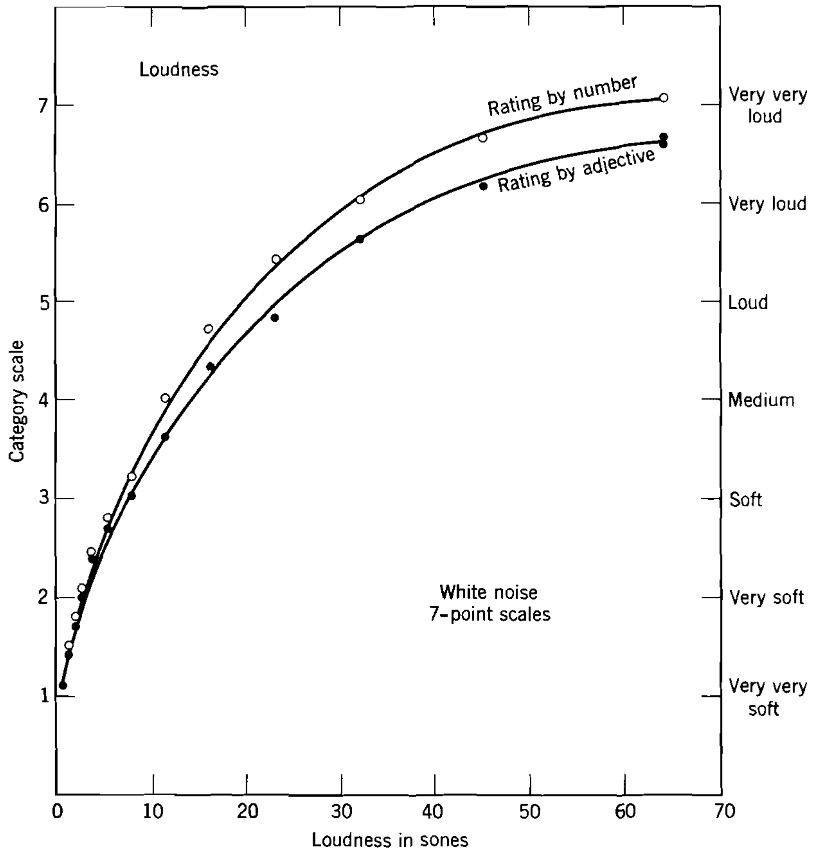

Fig. 5. Category rating scales for two different labelings of the responses. Adapted by permission from Stevens & Galanter (1957, p. 391).

is independent of the labeling (Stevens & Galanter, 1957). For example, scales computed from the first seven integers and from an equal number of adjectives are shown in Fig. 5. The similarity of the two functions is clear.

Varying the number of categories used has some effect, but it is small (Stevens & Galanter, 1957). The data shown in Fig. 6 compare $k = 3$ with $k = 100$.

There is considerable freedom in choosing instructions, and were they to affect the results appreciably the method would be judged poor. In general, however, the exact instructions used seem to have little effect as long as they ask the subject to make the intervals subjectively equal. The initial judgments seem to be somewhat influenced by the instructions, but if one permitted the subject to continue judging until his behavior stabilized, the functions would all be about the same (Stevens & Galanter, 1957). For this reason, most experimenters attempt to find and use instructions that cause subjects to achieve asymptotic stability rapidly. This result does, however, argue against averaging single judgments from a number of subjects.

The variables that have really important effects are those concerning the stimuli. The spacing of the stimuli along the physical continuum has noticeable consequences on the mean scale values because subjects tend to devote more categories to a stimulus interval as the density of stimuli in that interval is increased. It is as though the subjects were trying to spread out stimuli that are in fact close together, or, what is the same thing, to name the categories about equally often. This affects the apparent slope

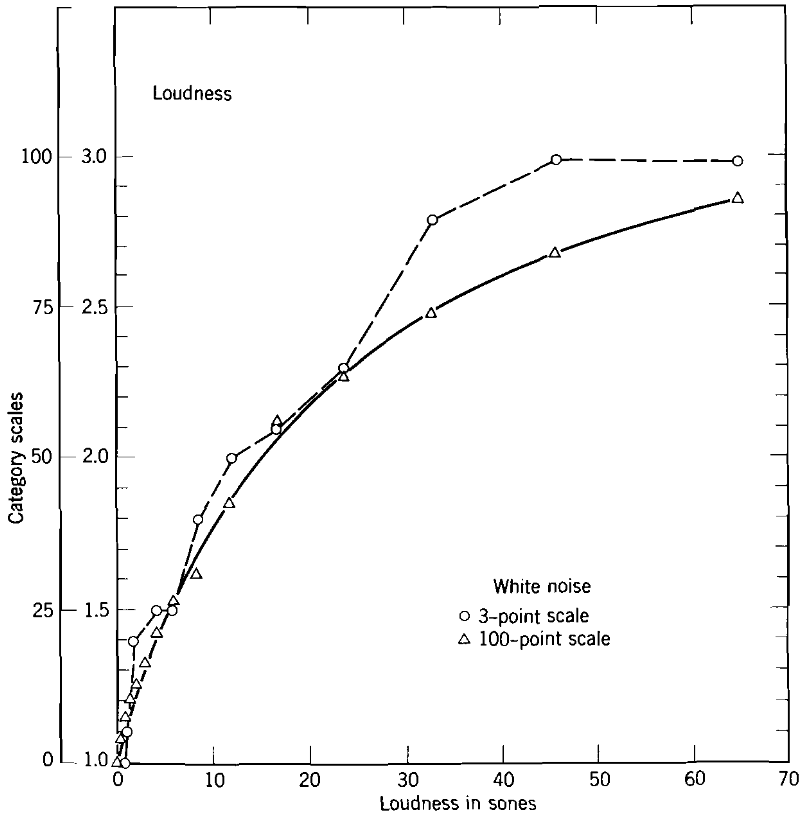

Fig. 6. Category scales of loudness with 3 and 100 categories. Adapted by permission from Stevens & Galanter (1957, p. 391).

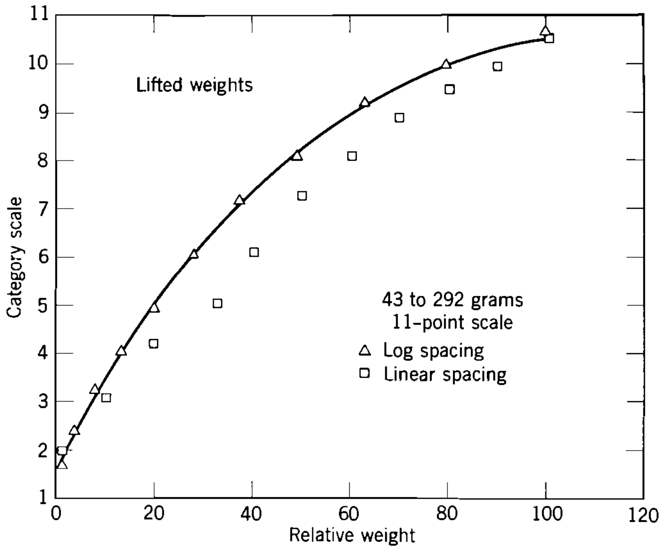

Fig. 7. Category scale of weight for two different stimulus spacings. Adapted by permission from Stevens & Galanter (1957, p. 384).

of the function: when the stimuli are closely packed, the function tends to be appreciably steeper in that region than when they are less dense (Stevens & Galanter, 1957). Examples are shown in Fig. 7. Essentially the same finding occurs if the spacing of the stimuli is held fixed and their presentation frequencies are varied. The function is steepened in regions of high presentation probability (Parducci, 1956). Together, these results suggest that the controlling variable is motivational, namely the relative use of the response categories.

Possibly related to this spacing effect is the so-called "anchoring effect." If a particular stimulus is selected as an anchor and is presented prior to every trial, then the function is always steeper in the vicinity of the anchor than when none is used (Michels & Doser, 1955). Alternatively, this has been interpreted as a purely stimulus effect, the anchor affecting the sensitivity of the subject in its neighborhood.

A full understanding of these effects cannot be expected until we have a sophisticated theory of category judgments. Unfortunately, what is now available is not fully satisfactory. Basically, the problem is to find a response theory which defines a scale of sensation that is invariant under the various experimental manipulations we have just described and does not depend upon an arbitrary, albeit conventional, labeling of the responses.

3.3 Successive Intervals and Related Scaling Models

The most widely known *model* for the analysis of category judgments is an adaptation of Thurstone's equation of comparative judgment for discrimination. Various versions have been discussed, the first by Saffir (1937) for what is known as successive intervals scaling and the most general by Torgerson (1954). The special cases that have been examined in detail are described by Torgerson (1958).

As before, each presentation s_i is assumed to result in a number on a subjective decision continuum, this number being a normally distributed random variable $\mathbf{S}_i$ with mean $\bar{s}_i$ and standard deviation σ_i. The subject's problem is assumed to be the assignment of the presentation to one of the k ordered response categories on the basis of this observation. The assumed decision rule is that the subject partitions the decision continuum into k intervals which are in one-to-one correspondence with the responses and that he responds r_j if and only if $\mathbf{S}_i$ lies in the jth interval. This partition is characterized by the $k - 1$ boundary values of the intervals ($-\infty$ and $+\infty$ need not be explicitly included as boundary values). The upper boundary point of interval j, $j < k$ is assumed to be a normally distributed random variable $\mathbf{T}_j$ with mean $\bar{t}_j$ and standard deviation τ_j.

The basic relative frequencies are assumed to estimate underlying probabilities $p(r_j \mid s_i)$. The cumulative

$$P(r_j \mid s_i) = \sum_{h=1}^{j} p(r_h \mid s_i)$$

is the probability that stimulus presentation s_i is assigned to one of the first j categories. By the decision rule, we see that

$$P(r_j \mid s_i) = \Pr(\mathbf{T}_j - \mathbf{S}_i \geqslant 0).$$

Paralleling the argument of Sec. 3.1 of Chapter 4, if $Z(j, i)$ is the normal deviate corresponding to $P(r_j \mid s_i)$ and if r_{ij} is the correlation between the two random variables $\mathbf{S}_i$ and $\mathbf{T}_j$, then

$$\bar{t}_j - \bar{s}_i = Z(j, i)(\sigma_i^2 + \tau_j^2 - 2r_{ij}\sigma_i\tau_j)^{1/2}. \tag{2}$$

This is known as the *equation* (or sometimes law) *of categorical judgment.*

The general model cannot be solved because there are $2(k + m - 2) + (k - 1)m$ unknowns (the $\bar{s}_i$, $\bar{t}_j$, σ_i, τ_j, r_{ij}) and only $(k - 1)m$ equations (not km because the last cumulative must be 1 for each stimulus). Various simplifying assumptions, similar to those for the equation of comparative judgment, have been explored and corresponding computational schemes have been worked out (see Torgerson, 1958). Before the general availability of high-speed computers, attention was largely confined to equal variance models, even though many workers suspected them to be wrong. Recently, a computer program for the case of zero correlations and unequal variances was described by Helm (1959).

By and large, the scale values found from this model are closely similar to those found simply by calculating the mean category number except that the calculated function has somewhat more curvature against the physical measure. This can be seen, for example, in Galanter and Messick's (1961) study of the loudness of white noise. They found that the scale values were approximately a logarithmic function of the energy level of the noise. Moreover, we suspect that this "processed category" scale is more invariant under changes in stimulus spacing, presentation probabilities, category labels, etc., than the mean category judgments, but we know of no research to prove it.

In an important special case of the equation of categorical judgment the category boundaries are assumed to be fixed, not random variables. This is known as the *successive intervals model*, and it has been carefully investigated by Adams and Messick (1958). Their main results are quoted by Suppes and Zinnes in Section 4.5 of Chapter 1. For this model, Eq. 2 reduces to

$$t_j - \bar{s}_i = Z(i,j)\sigma_i. \tag{3}$$

It is clear from Eq. 3 that

$$\begin{aligned} Z(i,j) &= \frac{t_j - \bar{s}_i}{\sigma_i} \\ &= \frac{t_j - \bar{s}_{i'} + \bar{s}_{i'} - \bar{s}_i}{\sigma_i} \\ &= \alpha(i,i')\,Z(i',j) + \beta(i,i'), \end{aligned} \tag{4}$$

where

$$\alpha(i,i') = \frac{\sigma_{i'}}{\sigma_i}$$

$$\beta(i,i') = \frac{\bar{s}_{i'} - \bar{s}_i}{\sigma_i}.$$

Thus the linear equation (4) is a necessary condition for the successive intervals model to hold; Adams and Messick also showed that it is a sufficient condition.

We note that there are $2m + k - 3$ unknowns in this model, which is not greater than $m(k - 1)$, the number of equations, when $k \geqslant 3$, which it always is.

3.4 A Choice Analysis

The other published models for category data (see Chapters 12 and 13 of Torgerson, 1958) all assume data from a group of subjects. A simultaneous analysis of responses and subjects based upon an assumed common scale is then performed. These methods belong to psychometrics, not psychophysics. Rather than go into them, we conclude this section by describing a simple choice model. No work has yet been done on the estimation of its parameters and, therefore, on its ability to account for data.

In essence, the idea is to collapse implicit responses into response categories just as was done in the analysis of the detection of an unknown stimulus (Sec. 9 of Chapter 3). Specifically, we suppose that underlying the observed category judgments are implicit recognition responses t_j which satisfy the choice model described in Sec. 1.2 of Chapter 3; the matrix of scale values is of the form

$$\begin{array}{c|ccccc|}
 & t_1 & \cdots\cdots & t_j & \cdots\cdots & t_m \\
\hline
s_1 & \eta(s_1, s_1)b_1 & \ldots & \eta(s_1, s_j)b_j & \ldots & \eta(s_1, s_m)b_m \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
s_i & \eta(s_i, s_1)b_1 & \ldots & \eta(s_i, s_j)b_j & \ldots & \eta(s_i, s_m)b_m \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
s_m & \eta(s_m, s_1)b_1 & \ldots & \eta(s_m, s_j)b_j & \ldots & \eta(s_m, s_m)b_m \\
\hline
\end{array}. \tag{5}$$

We assume that the subject's overt category responses are formed by partitioning the implicit responses into k classes in some unknown way. If we confine our attention to simply ordered stimuli and implicit responses, it seems reasonable to postulate partitions that can be defined in terms of $k - 1$ boundary points. We suppose that the set R_1 of implicit responses corresponding to category 1, that is, to response r_1, consists of all implicit responses $t_1, t_2 \ldots$ up to and including some last one, whose index we label r_1. The set R_2 of implicit responses corresponding to category 2 consists of the next implicit response after r_1 and all others up to and including a last one, whose index we label r_2; and so on. Thus the name of a response category and the index of the largest implicit response in that category have the same symbol.

Working again with cumulative response probabilities, we see from Eq. 5 that the fundamental equations are

$$\begin{aligned} P(r_j \mid s_i) &= \sum_{h=1}^{j} p(r_h \mid s_i) \\ &= \sum_{h=1}^{j} \sum_{l \in R_h} p(t_l \mid s_i) \\ &= \sum_{l=1}^{r_j} p(t_l \mid s_i) \\ &= \frac{\sum_{l=1}^{r_j} \eta(s_i, s_l) b_l}{\sum_{h=1}^{m} \eta(s_i, s_h) b_h} . \end{aligned}$$

The unknowns are the $k - 1$ category boundary indices r_j, the $m - 1$ implicit response biases b_h, and the $m(m - 1)$ stimulus parameters $\eta(s_i, s_j)$, a total of $(m + 1)(m - 1) + k - 1$ unknowns. There are only $m(k - 1)$ independent equations, and so, like the general Thurstonian model, the general choice model cannot be solved, even in principle.

Because we assumed that the stimuli were ordered, it is just as plausible here as it was in the study of similarity to suppose that the analogues of Assumptions 1, 2, and 3′ of Sec. 1.3 are satisfied. Then there are only $m - 1$ independent stimulus parameters, namely the $\eta(s_{i+1}, s_i)$ between adjacent pairs of stimuli. In that case the number of unknowns, $2m + k - 3$, does not exceed the number of equations provided that $k \geqslant 3$, which it always is. No workable scheme to find these unknowns is yet available.

Note that the unknown partition of the implicit responses into k classes has the same form as an identification function for a partial identification experiment, and so this choice analysis amounts to treating the category experiment as a recognition experiment of the partial identification variety in which the identification function is unknown. At the beginning of this chapter we suggested that ultimately this may be the way that all problems of this general type will be handled.

We observed earlier that the category methods are sensitive to the presentation probabilities, and, although we know of no relevant data, they are undoubtedly sensitive to payoffs. In terms of the foregoing model, these observed alterations in the response probabilities could correspond to adjustments in the response biases, b_h, in the category boundaries (i.e., the unknown identification function), or in both. It would be interesting to know which is affected. A reasonable conjecture is that the instructions fix the identification function and that the presentation probabilities and payoffs influence only the response biases in the underlying recognition model. Unfortunately, little can be done to answer these questions until we learn how to estimate the parameters, and that appears to be difficult.

4. MAGNITUDE ESTIMATION SCALES

4.1 The Method

Magnitude estimation and a number of allied methods evolved mainly from a program of research begun in the 1930's by S. S. Stevens to find better psychophysical scaling procedures than those of Fechner and Thurstone. One difficulty with the classical schemes is their reliance upon confusion among stimuli. They generate scales only for regions within which behavioral inconsistencies exist. If scales of wider range are desired, the local ones have to be pieced together in some fashion, usually by assuming that the subjective impression of a jnd is the same throughout the continuum or something nearly equivalent to that. This assumption Stevens doubted. A second difficulty is that these traditional methods at best yield interval scales, that is, they are unique only up to positive linear transformations, so that neither the zero nor the unit can be specified in a nonarbitrary way. For dimensions of intensity, such as sound intensity, it is difficult to believe that the subjective attribute, loudness, has an unspecified zero; there seems to be a reasonably well-specified level of stimulation below which there is no sensation and certainly negative loudnesses do not exist. Frequently, the threshold is chosen to be the zero, but this is an afterthought, not an integral part of the scaling model itself.

Among the methods that Stevens and others explored were fractionation and multiplication in which a stimulus is presented and the subject is asked to adjust a variable stimulus to a value that is either half or twice as loud. On the assumption that the subjective scale value of the stimulus chosen is indeed half or twice that of the one first presented, it was established empirically that the subjective scale would have to be approximately a power function of the usual physical measure of the stimulus, not the logarithmic function arising from Fechner's and Thurstone's models. Of course, the important and totally untested assumption of this model is the way in which the terms "one half" and "twice" in the instructions are assumed to be used by the subject in arriving at his judgments.

Having introduced "numbers" at all, it was not much of a leap to employ them in a much more massive way. In the resulting *method of magnitude estimation* the subject is instructed to assign a number to each stimulus presentation so that the numbers are proportional to the subjective magnitudes produced by the stimuli. Thus, for example, if one stimulus has been called 50 and another one seems subjectively one fifth as intense, it is to be called $10 = 50/5$. One stimulus is sometimes designated a standard and assigned a particular response value by the experimenter; usually 1, 10, or 100 is chosen so that fractional computations are easier for subjects. As early as 1956, however, S. S. Stevens showed that by not using a standard one can avoid certain local perturbations. Other methods that give essentially identical results have been described by S. S. Stevens (1958a). We shall confine our attention to magnitude estimation on the reasonable assumption of first-order experimental equivalence among these methods.

In some ways magnitude estimation and recognition experiments are alike. In both, the subject is more or less explicitly urged to make a unique response to each different stimulus presentation, that is, to act as if there were a one-to-one correspondence between stimuli and responses. A major difference between the two experiments is the size of the presentation and response sets. As far as the subject knows in a magnitude estimation experiment, any possible stimulus magnitude may be presented, although in practice the experimenter uses only relatively few. The subject's responses are restricted by the instructions only to the positive real numbers, although subjects seldom if ever use numbers other than integers and simple fractions.

As in recognition experiments, subjects do not consistently assign the same number to a particular stimulus. The inconsistencies are large enough that, in our opinion, they cannot be dismissed as analogous to the errors of measurement familiar in physics. The standard deviation[3] of the responses to a particular stimulus is somewhere in the neighborhood of 20 to 40 per cent of the mean response value, whereas, in good physical measurement the errors are usually reduced to less, often considerably less, than one per cent of the mean. The variability of magnitude estimation data appears to be due mostly to the subject, not to our equipment or

[3] There is a problem of conventional usage here. The subject's responses are not numbers but rather utterances or marks that conventionally name numbers. These names can no more be manipulated as numbers than can, say, responses that are color names. Numerical manipulations do, however, make sense if the responses are converted into random variables by establishing a one-to-one correspondence between the possible responses and a set of real numbers. We can then speak of the expectation of various functions of these random variables. Because the obvious one-to-one correspondence, namely, the assignment to a response of the number usually designated by the utterance made, is always used when analyzing magnitude estimation data, it is conventional to drop any reference to this correspondence and to treat the responses as if they were actually numbers. We shall follow this convention throughout the rest of this section.

recording procedures, and so it is an inherent part of the phenomena under study.

These observations suggest that the process can be effectively described only by a probabilistic model. But, in contrast to the models up to this point, we cannot hope to estimate in detail the relevant conditional probabilities, namely, the probabilities that particular numbers are emitted in response to a given stimulus presentation. There are simply too many possible responses to make that practical. We may, of course, postulate that such probabilities exist, but in tests of any such model we shall have to content ourselves with summary properties that can be easily estimated from data. Such a probabilistic model is described in Sec. 4.4, but as background we need to know more about experimental practice.

When studying a single subject, each of several (usually 10 to 20) stimuli are presented a number of times, and some "average" value of the responses to each is taken to be its "magnitude scale" value. The median, mean, and geometric mean have all been used at one time or another. Because the mean is the unbiased estimate of the expected value of the response, one might expect it to be favored; however, the data are almost always plotted in logarithmic coordinates, and, because the distribution of responses is approximately symmetrical and the geometric mean is the unbiased estimate of the quantity plotted, Stevens has recommended that it be used.

Most of the published data have not, however, been for single subjects. Rather, one or two responses per stimulus have been obtained from each subject, and an "average" over subjects is taken as the magnitude scale. The defense for averaging over subjects is that we are interested in central tendencies, not individual differences, and so the typical scale is what we want. Moreover, there are very practical engineering reasons for having standardized social scales for certain important dimensions such as loudness. Without disputing either point, we hope that more data for individual subjects will be published, because until it is decided just what it is that is invariant from subject to subject it will be difficult to be sure just what sort of averaging is permissible.

However the averaging is done, the resulting magnitude scale values are plotted against a physical measure of the stimuli; usually, both scales are shown in logarithmic coordinates. For continua involving changes of intensity, or what Stevens and Galanter (1957) called prothetic ones, the magnitude scale ψ is to a fair approximation a power function of the physical energy s of the stimulus, that is, there are constants α and β such that for stimulus values not too near threshold

$$\psi(s) = \alpha s^{\beta},$$

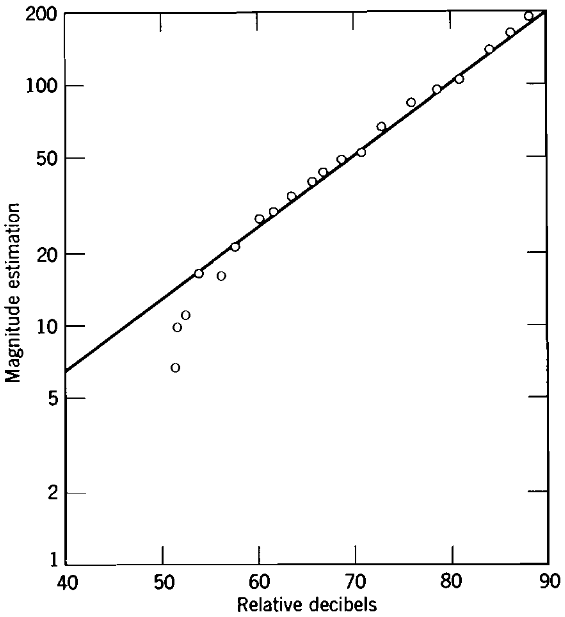

Fig. 8. Magnitude estimation judgments of loudness plotted in log-log coordinates. The straight line represents the power relation of loudness to intensity that has been accepted by the International Standards organization; the exponent is 0.3. Adapted by permission from Galanter & Messick (1961, p. 366).

or in logarithmic coordinates the relation is approximately a straight line with slope β (see Fig. 8). The departure from a straight line for small stimulus values is discussed later.

For each modality the exponent β is a reproducible quantity, not an unestimable parameter arbitrarily selected by the experimenter, whereas the constant α is a free parameter, whose value depends upon the units of both the physical and response scales. A listing of typical exponents for several different modalities is given in Table 1.

4.2 The Psychophysical Law

Historically, the relation between a measure of the subjective magnitude of sensation and a physical measure of the corresponding physical variable has been called the *psychophysical law*. There have been but two major contenders for the form of this relation. The first to appear, and the more dominant one throughout the history of psychophysics, was Fechner's logarithmic function, which we discussed at some length in Sec. 2 of

Chapter 4. Various modifications of his procedures and theory have evolved over the years, but with the exception of the relatively unsatisfactory mean category scale, all of them have rested upon some assumption that permits one to piece together the function from relatively local inconsistencies in the data. Neither a direct measurement of the subjective scale nor a satisfactory test of these assumptions has ever been suggested.

Table 1 Power Function Exponents of Magnitude Scales for Various Continua

Attribute	Exponent	Stimulus Conditions
Loudness	0.30	Binaural, 1000-cps tone, measured in energy units
Loudness	0.27	Monaural, 1000-cps tone, measured in energy units
Brightness	0.33	5° target, dark-adapted eye
Vibration	0.95	60 cps, finger
Vibration	0.6	250 cps, finger
Duration	1.1	White noise stimulus
Heaviness	1.45	Lifted weights
Electric shock	3.5	60 cps, through finger

Adapted from S. S. Stevens (1961b). Each exponent was determined by averaging data from at least ten subjects.

The alternative relation, the power function, was early suggested as a substitute to Fechner's proposal; it was briefly debated and then was forgotten for many decades until Stevens developed the method of magnitude estimation and discovered that it, not the logarithm, was consistent with his data. Buttressed by extensive experimentation, Stevens has argued that the power function is the correct form for the psychophysical law.

It is not our business here to recount these experimental studies nor to recapitulate all of Stevens' arguments; detailed summaries can be found in Stevens (1957, 1960, 1961a,b), where references are given to the numerous relevant experimental papers. Suffice it to say that he has repeatedly shown for a variety of prothetic continua that the magnitude scale is to a good approximation a power function of the physical scale and that he has created an elaborate network of consistent, interrelated results matched neither in detail nor in scope by those who adhere to the logarithmic form for the psychophysical law.

If this is so—and we suspect that most psychophysicists will agree that Stevens has amassed an impressive pattern of results—can there be any question about the form of the psychophysical law? His methods are direct, they do not involve Fechner's untested—and quite possibly untestable—assumption about the relation between magnitude and variability, and they have led to a structure of empirical relations which has few rivals for scope in psychophysics. Can there be doubt that the power function is the psychophysical law? Yet there is.

There are many detailed questions, but in our view the central one is: what meaning can we attach to an average of the numerical responses that a subject emits to a stimulus? Is it defensible or not to treat this as a measure of subjective sensation? Because this question seems so essential in resolving the debate between Stevens and his critics and because it is just the sort of question to which a mathematical theory might be brought to bear, we shall focus the rest of our discussion upon it.

4.3 The Invariance of the Scale

Averaging of one sort or another is certainly a legitimate way to condense and summarize one's observations, but that does not necessarily justify treating these numbers as a measure of anything—in particular, as scale values or as estimates of scale values. For example, in Sec. 3.2 we criticized an analogous averaging procedure in category scaling, and we might be expected to apply the same objections here with only a slight rewording. We shall not, however, for this reason. The trouble there was that the category numbers were assigned by the experimenter in a way that is arbitrary except for their ordering. Assuming that the subject's responses are independent of the labeling used, the experimenter can generate any monotonic function he wishes to by his choice of numbers to relate the mean category judgments to the physical scale. Magnitude estimation differs in that the subject, not the experimenter, chooses the numbers, and so they cannot be considered arbitrary in the same sense; it requires empirical evidence to know just how arbitrary they are.

The essence of the matter is probably the degree of arbitrariness of the responses, not the fact that they are utterances of number names. To be sure, one of the first objections to magnitude estimation was the numerical responses; there was an uneasy feeling that they must reflect more about the subject's number habits than about his sensory processes. To counter this view, Stevens (1959) argued in the following way that numerical responses are not essential to his results.

Let stimuli from one dimension, such as sound intensity, be presented, and during or just after each presentation let the subject adjust stimulation from another modality, such as skin vibration, until it "matches" the first in intensity. (The concept of a cross-modality match is left undefined, as are, of course, the matching operations basic to most physical measurement. The difference is that physical matching usually involves two stimuli on the same physical dimension, not from two different ones, and it is generally conceded that the former is a far simpler judgment.) Let us suppose that values of the two physical scales are s and t, that the two subjective scales are power functions

$$\psi(s) = \alpha s^{\beta} \quad \text{and} \quad \psi^*(t) = \alpha^* t^{\beta^*},$$

and that matching is defined as meaning equal subjective values, that is, s is the matching stimulus to t if and only if

$$\psi(s) = \psi^*(t).$$

It follows immediately that

$$s = \left(\frac{\alpha^*}{\alpha}\right)^{1/\beta} t^{\beta^*/\beta}.$$

These assumptions imply that the matching relation is a power function with the exponent β^*/β. Thus, if the magnitude scales represent subjective sensation, we predict not only that the matching data will follow a power function, but we also predict the value of the exponent. Both predictions have been confirmed in a variety of cases (see Stevens 1959, 1961b); sample data for vibration and sound intensity are shown in Fig. 9, where the theoretical line is a parameter-free prediction based upon the magnitude estimation data for each modality separately. The confirmation of the exponent is impressive. Comparable data for individual subjects have not been published.

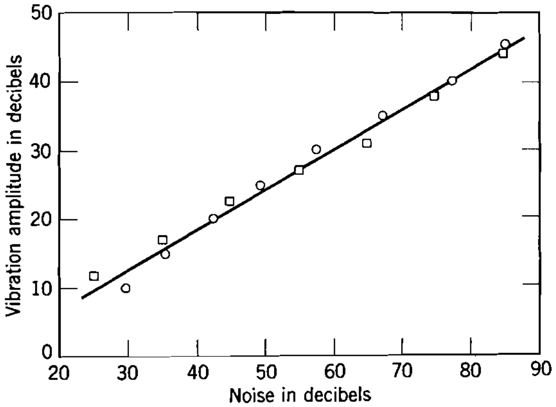

Fig. 9. The observed matching relation between the noise intensity and vibration amplitude. The theoretical line is predicted from magnitude estimation data on each modality separately. Adapted by permission from S. S. Stevens (1959, p. 207).

Critics have argued that this outcome of cross-modality matching is not really an argument supporting the power law, because if

$$\psi(s) = a \log \frac{s}{b} \quad \text{and} \quad \psi^*(t) = a^* \log \frac{t}{b^*},$$

then

$$a \log \frac{s}{b} = a^* \log \frac{t}{b^*},$$

or, taking exponentials,

$$s = ct^{a^*/a} \quad \text{where} \quad c = b\left(\frac{1}{b^*}\right)^{a^*/a}.$$

Thus both the power function and logarithmic hypotheses predict a power relation for the matching data, and so, it is argued, either is equally acceptable. This overlooks the fact that the two hypotheses differ in what they say about the exponent of the matching relation. If the magnitude scales are power functions, then the exponent of the matching data is given as the ratio of the estimable exponents of the magnitude functions; whereas, if they are logarithmic, the exponent is nothing but the ratio of the arbitrary units of the two scales, and so it is not predicted. The fact that the obtained exponents are well predicted by those from magnitude estimation leads us to favor the power function over the logarithm.

This last argument, however, somewhat prejudges the issue in question by assuming that we know the number of free parameters, and that is what is uncertain. Stevens has frequently referred to magnitude estimation as a "ratio scaling method" (e.g., Stevens, 1957, 1961b), which in this context is an unhappily ambiguous phrase. On the one hand, it might be purely descriptive of his method, referring to the fact that the subjects are asked to use numbers so that subjective ratios are preserved. On the other hand, it might be and usually is interpreted to mean that the resulting scale is technically a ratio scale, that is, it is completely specified except for its unit (see Chapter 1). No one can object to the descriptive use of the phrase except to the extent that it automatically suggests the second, much more significant, meaning. This extent seems to be great.

As Suppes and Zinnes point out in Chapter 1, the decision about the type of scale—ratio, interval, ordinal, etc.—is ultimately a theoretical one. One states certain axioms, for example, about some primitive concatenation operation and some binary relation corresponding to the judgments made. If these axioms are not empirically falsified in a few tests, they are assumed to be generally true. One then shows mathematically that a certain numerical representation exists which is isomorphic to the axiom system, and the scale type is determined by showing the group of transformations that characterizes all isomorphic representations into the same numerical system. If that is what is meant by constructing a scale of a particular type, then it is clear that we can be certain neither that the numbers obtained by magnitude estimation form a scale in this sense nor, if they do, what their scale type is until an explicit measurement theory is stated.

One can, however, hardly expect the empirical scientist to discard a method that seems to give regular and reproducible results just because no satisfactory theory exists to account for them. Rather, he will attempt to show by various experimental manipulations that the magnitude scale for a given physical dimension appears to have the invariances attributed to it. This Stevens has done. See Stevens (1960, 1961a,b) for summaries of this work. We examine several aspects that seem particularly relevant to the invariance question.

It was early noted that in log-log coordinates the functions are not quite straight lines and that they rise more rapidly for low stimulus levels than for the medium and high ones, where they are straight. This can be seen in Fig. 8. Moreover, if the subject's threshold is artificially inflated by introducing a masking background, the effect becomes more pronounced, as shown in Fig. 10. This lack of invariance in the form of the function can be interpreted either as showing that the magnitude scale simply is not a ratio scale or as showing that the relevant variables are not being used. The simple power relation states that the scale value approaches its zero as the physical variable approaches its zero, but we know perfectly well that stimuli cannot be detected for energy levels up to what is called the subject's threshold—absolute or artificial, as the case may be. This suggests that one of the two scales is wrong. One possibility suggested by a number of writers is to modify the equation to read

$$\psi(s) = \alpha(s - \gamma)^{\beta}, \qquad s \geqslant \gamma > 0, \tag{6}$$

where α is again an unspecified parameter (namely, the unit of the magnitude scale) and β and γ are estimable parameters which supposedly depend upon the conditions of stimulation—the nature of the stimulus presentations, the background, etc.

An alternative, suggested by McGill (1960), is to add the new parameter to the scale values, that is,

$$\psi(s) = \alpha(s^{\beta} - \delta), \qquad s \geqslant \delta^{1/\beta} > 0. \tag{7}$$

The variability of magnitude estimation data is such that both functions fit equally well, and so the decision will probably have to be reached indirectly. An example of such an attempt is the fairly elaborate argument

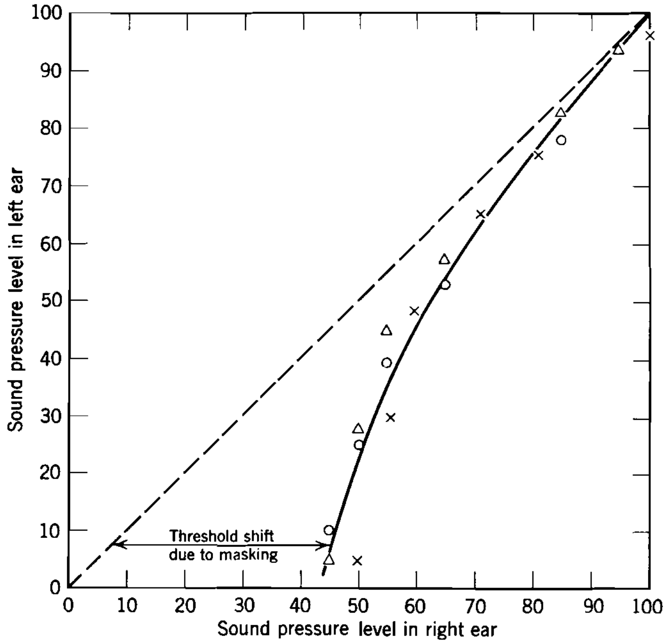

Fig. 10. The apparent increase of the threshold produced by a masking background. Adapted by permission from Stevens (1958, p. 519); the data were originally reported by Steinberg and Gardner (1937).

given by Galanter and Messick (1961), based in part upon some more or less philosophical considerations of Luce (1959), which suggests that Eq. 6 may be more appropriate than Eq. 7. In our opinion, however, the matter remains open.

It is now generally conceded that no matter where the "threshold" constant, γ or δ, is placed, the scale should be viewed as involving two estimated parameters—the threshold constant and the exponent β—and one unestimable parameter—the unit α—rather than treated as something weaker than a ratio scale. By weaker, we mean an interval scale or any other type in which there are two or more unestimable parameters. Incidentally, the previous arguments about the matching data are unaffected by the addition of such a "threshold" parameter.

Other empirical attempts to show that it is reasonable to treat the magnitude values as numbers on a ratio scale have involved showing that the form and the estimates of β and γ are invariant under modifications of the wording of the instructions, use of different number assignments to the standard stimulus, different locations of the standard, and variations in the number and spacing of the stimuli (Stevens & Galanter, 1957; J. C. Stevens, 1958). In all cases the data have been interpreted as supporting, to a first approximation, the desired invariance.

In spite of all this favorable experimental evidence, we still doubt that the magnitude scale is completely specified except for its unit. To anticipate the coming argument, we suspect that a psychophysical ratio scale exists which under certain conditions is well estimated by the magnitude scale. Nevertheless, just as in the rest of psychophysics, we suspect that a subject's responses, and therefore the magnitude scale, are some composite of his sensations and of other factors which, for lack of a better term, we call motivational. Our problem, therefore, is to attempt to construct a theory that makes these dependencies explicit and then to ask under what conditions is it reasonable to view the magnitude scale as a satisfactory estimate of the underlying, invariant psychophysical scale.

4.4 A Probabilistic Response Theory

As pointed out earlier, a magnitude estimation or a cross-modality matching experiment is much like the complete identification experiments discussed in Chapter 3. If we label stimuli by their physical magnitudes, then for a continuous dimension the set S of stimuli can be identified with the set of positive real numbers. Similarly, whether the responses are actual numbers or the physical measures of a matching variable, R can also be treated as the positive real numbers. Aside from the fact that S and R are no longer small finite sets, a magnitude estimation experiment also differs from a complete identification experiment in that no identification function ι is specified by the experimenter. Such a function we shall assume is induced in the subject by the instructions, and that function is just what one hopes to discover from the data.

This point is crucial in the development of a theory for magnitude estimation. Recall that in an identification experiment the identification function $\iota: R \rightarrow S$ is established by the experimenter and is communicated to the subject by the instructions and information feedback. Such an experiment is not considered under proper control until the identification function is specified. If, however, we are dealing with a situation in which we believe that the subject has what amounts to his own identification function and if we want to know what it is, then introducing our own arbitrary one would only help to conceal the unknown one of interest. Rather, we must let the subject be free to reveal the one he has. Magnitude estimation is one way that has been proposed for him to do this.

On the surface, there seems to be an inconsistency, for now we are saying that magnitude estimation is designed to get at the unknown identification function, whereas earlier we suggested that the unknown function is the psychophysical scale. If, however, we postulate that $\iota: R \rightarrow S$ is a strictly monotonic increasing function, then its inverse $\psi: S \rightarrow R$ exists, and so determining one is equivalent to determining the other. We shall suppose that ψ is the psychophysical scale, whereas its inverse ι is the identification function.

Because we know that subjects often fail to give the same response when a stimulus is repeated, more than ψ must be involved in relating responses to stimuli. Previously, we have had to invoke some notion both of response bias and of stimulus generalization to account for psychophysical data, and so we do it again using a continuous analogue of the choice model for complete identification experiments (Sec. 1.2, Chapter 3).

Let $p(r \mid s)$ denote the conditional probability density of response r to stimulus s, let b be a real-valued function defined over R, which represents the response bias, and let $\eta(s, t)$ denote a measure of generalization from stimulus s to stimulus t. The model postulates that

$$p(r \mid s) = \frac{\eta[s, \iota(r)]\, b(r)}{\int_0^\infty \eta[s, \iota(x)]\, b(x)\, dx}. \tag{8}$$

Observe that if we define a real-valued function ζ over $R \times R$ in terms of η, namely

$$\zeta(x, y) = \eta[\iota(x), \iota(y)], \quad x, y \in R,$$

then, by a simple substitution and taking into account that $\psi = \iota^{-1}$, Eq. 8 can be rewritten as

$$p(r \mid s) = \frac{\zeta[\psi(s), r]\, b(r)}{\int_0^\infty \zeta[\psi(s), x]\, b(x)\, dx}. \tag{9}$$

Thus it is immaterial whether we view the generalization as over stimuli or over responses, but for certain later computations it is more convenient to use the second form.

In words, Eq. 8 assumes that when stimulus s is presented it has some chance, which is proportional to $\eta(s, t)$, of seeming like stimulus t, and the subject responds to t according to his psychophysical function ψ. Thus the response is $r = \psi(t)$. In Eq. 9 s leads to the sensation $\psi(s)$, but because of response generalization the response r is emitted with some probability proportional to $\zeta[\psi(s), r]$. Overlying this purely psychophysical structure is a response bias $b(r)$ which differentially influences the responses that occur. The crude, unnormalized measure of the strength of connection between stimulus s and response r is simply $\eta[s, \iota(r)]\, b(r)$ or, equivalently, $\zeta[\psi(s), r]\, b(r)$. The total measure is $\int_0^\infty \zeta[\psi(s), x]\, b(x)\, dx$, so that dividing the measure of the strength of connection by the total measure yields a probability, just as in the discrete choice models.

(For those familiar with Stieltjes integrals, it should be noted that Eqs. 8 and 9 should properly be written as integrals with respect to a cumulative bias B. When that is done, discrete choice models are simply special cases in which the cumulative bias is a step-function.)

Three general comments about this model are in order. First, although we have viewed it as a continuous generalization of the choice theory for recognition experiments, it is also much like the response mechanism that Thurstone postulated to explain discrimination data and Tanner used in his analysis of detection and recognition data. The principal differences are the multiplicative biasing function and the fact that the generalization function is not assumed to be normal. These differences in the continuous case are minor compared with those that develop when the models are applied to finite stimulus presentation and response sets. There the choice theory involves only those values of the generalization function for the specific stimuli employed, whereas in the Thurstonian theories whole sets of stimuli, or their corresponding responses, are treated as equivalent, and integrals of the generalization measure over these sets are treated as the needed discrete probabilities.

Second, the three functions $\psi = \iota^{-1}$, η or ζ, and b, which enter into Eqs. 8 and 9, have no necessary relation to one another. In particular, the measure of generalization, η or ζ, which is of paramount importance in the choice theory analyses of detection, recognition, and discrimination, need have no particular connection with the psychophysical scale ψ, which characterizes how sensation grows with stimulus energy. This point has been repeatedly emphasized by Stevens (e.g., 1961a, p. 83) in his criticisms of the classic attempts to derive the psychophysical function from discrimination data (Sec. 2, Chapter 4). Shortly, however, we shall see certain theoretical reasons why, in a sense, both points of view may be correct and why it has proved so difficult using only confusion data to disentangle the psychophysical and generalization functions.

Third, a formal analogue of the asymptotic learning argument given in Sec. 1.2 of Chapter 3 yields Eq. 8. Of course, the discrete probabilities of the model for complete identification experiments must be replaced by probability densities. It seems doubtful, however, that this learning model can be taken as a serious argument for the continuous choice model because no payoffs are used in magnitude estimation experiments. It may, however, suggest a way of analyzing data in which payoffs are used.

As mentioned earlier, there is little hope of estimating $p(r \mid s)$ from data, but various of its parameters can be estimated. Of particular relevance to magnitude estimation are the expected response and the "geometric expected response," that is, the exponential of the expectation of the logarithm of the response. These are defined by

$$\begin{aligned}\psi_m(s) &= E(r \mid s) \\ &= \int_0^\infty r p(r \mid s)\, dr \\ &= \frac{\int_0^\infty r\zeta[\psi(s), r]\, b(r)\, dr}{\int_0^\infty \zeta[\psi(s), r]\, b(r)\, dr}, \end{aligned} \tag{10}$$

and

$$\begin{aligned}\psi_g(s) &= \exp E(\log r \mid s) \\ &= \exp \int_0^\infty (\log r)\, p(r \mid s)\, dr \\ &= \exp \frac{\int_0^\infty (\log r)\zeta[\psi(s), r]\, b(r)\, dr}{\int_0^\infty \zeta[\psi(s), r]\, b(r)\, dr}, \end{aligned} \tag{11}$$

respectively.

Assuming that the model is correct, the most important question is: when is the theoretical magnitude scale ψ_m or ψ_g approximately proportional to the psychophysical function ψ? It is evident that we can choose ζ and b so that they are quite different. In Theorem 2 we state one set of sufficient conditions leading to proportionality, but first we show two ways of stating one of the conditions.

Lemma 1. *Suppose that ζ has the property that there exist positive continuous functions f and g such that for all $x, y, z > 0$, $\zeta(xz, yz) = f(x, y)\, g(z)$; then there exists a constant γ such that $\zeta(x, y) = x^\gamma \zeta(1, y/x)$. Conversely, if h is any positive, continuous function, then $\zeta(x, y) = x^\gamma h(y/x)$ has the foregoing property.*

PROOF. By setting $z = 1$, we see that $f(x, y) = \zeta(x, y)/g(1)$. Using this and the hypothesis three times, we obtain

$$\begin{aligned}\zeta\left(1, \frac{yz}{xz}\right) \frac{g(xz)}{g(1)} &= \zeta(xz, yz) \\ &= \zeta(x, y) \frac{g(z)}{g(1)} \\ &= \zeta\left(1, \frac{y}{x}\right) g(x) \frac{g(z)}{g(1)^2}.\end{aligned}$$

Setting $u(x) = g(x)/g(1)$ and dividing by $\zeta(1, y/x)$, we get $u(xz) = u(x)\,u(z)$. Because g is positive and continuous, so is u; hence the functional equation has the solution $u(x) = x^{\gamma}$ for some constant γ. Thus $g(x) = g(1)x^{\gamma}$. Substituting,

$$\begin{aligned}\zeta(x, y) &= f\left(1, \frac{y}{x}\right) g(x)\\ &= \zeta\left(1, \frac{y}{x}\right) x^{\gamma}.\end{aligned}$$

Conversely, if $\zeta(x, y) = h(y/x)x^{\gamma}$, then

$$\begin{aligned}\zeta(xz, yz) &= h\left(\frac{yz}{xz}\right) (xz)^{\gamma}\\ &= h\left(\frac{y}{x}\right) x^{\gamma}z^{\gamma}\\ &= \zeta(x, y)z^{\gamma}.\end{aligned}$$

Theorem 2. *If ζ has the property that $\zeta(x, y) = x^{\gamma}\zeta(1, y/x)$ and if $b(r) = br^{c}$ for all $r \in R$, then $\psi_m(s) = \mu\psi(s)$ and $\psi_g(s) = \mu_g\psi(s)$ for all $s \in S$, where μ and μ_g are, respectively, the mean and geometric mean of $r^{c}\zeta(1, r)\Big/\int_0^{\infty} x^{c}\zeta(1, x)\, dx$.*

PROOF. We prove this only for the mean; the other case is similar. If we let $x = r/\psi(s)$ and substitute our assumptions in Eq. 10, we obtain

$$\begin{aligned}\psi_m(s) &= \frac{\int_0^{\infty} \psi(s)x\zeta[\psi(s), \psi(s)x]bx^{c}\, \psi(s)^{c}\, \psi(s)\, dx}{\int_0^{\infty} \zeta[\psi(s), \psi(s)x]bx^{c}\, \psi(s)^{c}\, \psi(s)\, dx}\\ &= \psi(s)\frac{\int_0^{\infty} x^{1+c}\, \psi(s)^{\gamma}\zeta(1, x)\, dx}{\int_0^{\infty} x^{c}\, \psi(s)^{\gamma}\zeta(1, x)\, dx}\\ &= \psi(s)\mu,\end{aligned}$$

where

$$\mu = \frac{\int_0^{\infty} x^{1+c}\zeta(1, x)\, dx}{\int_0^{\infty} x^{c}\zeta(1, x)\, dx}.$$

The conclusion, then, is that both the expected response and the geometric expected response are proportional to the psychophysical function, provided that the response bias is a power function and that the generalization function has a particular form, the simplest case of which (i.e., $\gamma = 0$) postulates that generalization depends upon the ratio of the psychophysical function values of the two stimuli. Note that this conclusion is completely independent of the form of the psychophysical function ψ; therefore, if we can convince ourselves that the two assumptions of Theorem 2 are met in a magnitude estimation or matching experiment, then the observed magnitude scale estimates the underlying psychophysical function—which is what we want to measure.

In the next two sections we turn to questions about the mathematical form of the generalization and psychophysical functions. Some of the results about the generalization function appear to be helpful in deciding whether the conditions of Theorem 2 are met for a given set of data.

4.5 Form of the Generalization Function

To show that the expected response is proportional to the psychophysical function, we found it necessary to constrain the generalization function. This constraint is, however, quite different from those we seemed to need in analyzing detection and recognition experiments (Chapter 3). There we assumed that the negative logarithm of the generalization function has the properties of a distance function. Moreover, when the stimuli differ on only one physical dimension, we assumed that the distance measure was additive. Because these postulates have received some indirect support, it seems worthwhile to find out what they imply when added to the present constraint. The answer is given in the following theorem

Theorem 3. *If the generalization function ζ is such that*

1. $\zeta(x, y) = \zeta(1, y/x)x^{\gamma}$, *for all* $x, y > 0$,
2. $\zeta(x, z) = \zeta(x, y)\,\zeta(y, z)$, *for all* x, y, z *for which either* $x \geqslant y \geqslant z$ *or* $x \leqslant y \leqslant z$,
3. $\zeta(x, y) = \zeta(y, x)$, *for all* $x, y > 0$, *and*
4. ζ *is continuous in each of its arguments,*

then there exists a constant δ such that

$$\zeta(x, y) = \begin{cases} \left(\dfrac{x}{y}\right)^{\delta} & \text{if } x \leqslant y \\ \left(\dfrac{x}{y}\right)^{-\delta} & \text{if } x \geqslant y. \end{cases}$$

PROOF. By condition 2, $\zeta(x, y) = \zeta(x, x)\,\zeta(x, y)$, so $\zeta(x, x) = 1$. Using this and condition 1, $\zeta(x, x) = 1 = \zeta(1, 1)x^{\gamma}$, hence $\gamma = 0$. If $x, y \geqslant 1$, then conditions 1 and 2 imply

$$\begin{aligned}\zeta(1, x)\,\zeta(1, y) &= \zeta(1, x)\,\zeta(x, xy)\\ &= \zeta(1, xy).\end{aligned}$$

Because, by (4), $\zeta(1, x)$ is continuous, this functional equation is known to have the solution $\zeta(1, x) = x^{-\delta}$ for some δ. A similar argument holds when $x, y \leqslant 1$, leading to $\zeta(1, x) = x^{\epsilon}$, for some ϵ. It is easy to see that condition 3 implies $\epsilon = \delta$, thus proving the theorem.

We examine next how one might study the form of the generalization function empirically. If we assume that there is no response bias, that is, $c = 0$, and that the generalization function depends only upon response ratios, then by Theorem 2 we see that

$$\frac{r}{\psi_m(s)} = \frac{1}{\mu}\frac{r}{\psi(s)}.$$

Thus the distribution ζ^* of $r/\psi_m(s)$ is simply the distribution ζ of $r/\psi(s)$, except that the independent variable is stretched by a factor μ, that is, $\zeta^*(x) = \mu\zeta(\mu x)$. So, if we estimate $\psi_m(s)$ from the mean empirical response curve, then we can develop the empirical frequency distribution corresponding to ζ^*, which except for a constant multiplicative factor is the generalization function.

If we wish to test the hypothesis that the generalization function has the form derived in Theorem 3, we can use a χ^2 test once δ is estimated. At the moment we have only *ad hoc* techniques for estimating δ under the assumption of no response bias. Because the mean of ζ^* is easily seen to be 1 in this case, it cannot be used to estimate δ, but either the median or variance can be. Let M and M^* denote, respectively, the medians of ζ and ζ^*; then it is clear that $M^* = M/\mu$. For the generalization function of Theorem 3, if $\delta > 2$, the mean μ is given by

$$\begin{aligned}\mu &= \frac{\int_0^\infty x\,\zeta(x)\,dx}{\int_0^\infty \zeta(x)\,dx}\\ &= \frac{\int_0^1 x^{\delta+1}\,dx + \int_1^\infty x^{1-\delta}\,dx}{\int_0^1 x^{\delta}\,dx + \int_1^\infty x^{-\delta}\,dx}\\ &= \frac{\delta^2 - 1}{\delta^2 - 4}.\end{aligned}$$

The median is defined by

$$\frac{1}{2} = \frac{\int_0^M \zeta(x)\,dx}{\int_0^\infty \zeta(x)\,dx}$$

$$= \frac{\int_0^1 x^\delta\,dx + \int_1^M x^{-\delta}\,dx}{\int_0^1 x^\delta\,dx + \int_1^\infty x^{-\delta}\,dx}$$

$$= \frac{\delta - 1}{2\delta} - \frac{\delta + 1}{2\delta}\,[M^{(1-\delta)} - 1].$$

Solving,

$$M = \left(\frac{\delta + 1}{\delta}\right)^{1/(\delta-1)}.$$

Thus,

$$M^* = \left(\frac{\delta + 1}{\delta}\right)^{1/(\delta-1)} \left(\frac{\delta^2 - 4}{\delta^2 - 1}\right). \qquad (12)$$

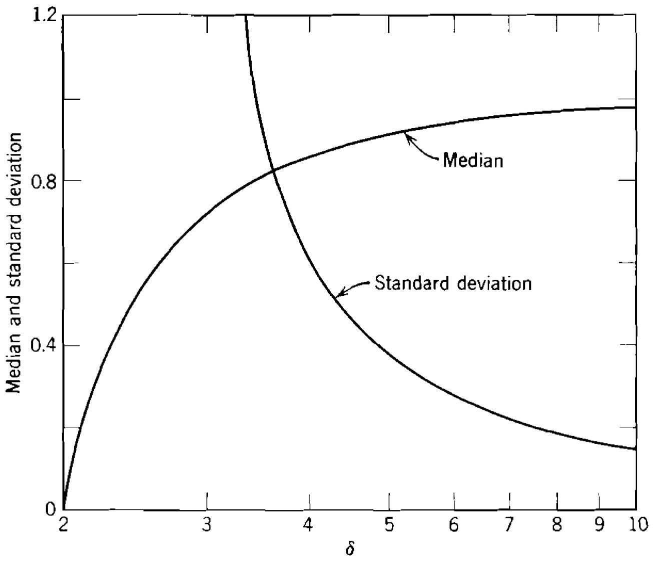

Fig. 11. The standard deviation and median of the generalization function

$$\zeta(x) = \begin{cases} \dfrac{\delta^2 - 1}{2\delta}\,x^\delta & \text{if } 0 < x < 1 \\[2ex] \dfrac{\delta^2 - 1}{2\delta}\,x^{-\delta} & \text{if } 1 < x. \end{cases}$$

So an empirical estimate of the median provides an estimate of δ. Equation 12 relating δ to M^* is plotted in Fig. 11.

Similarly, if $c = 0$ and $\delta > 3$,

$$\text{var}\,(\zeta^*) = \frac{\text{var}(\zeta)}{\mu^2}$$

and

$$\text{var}\,(\zeta) = \frac{\delta^2 - 1}{\delta^2 - 9} - \mu^2,$$

hence

$$\text{var}\,(\zeta^*) = \frac{(\delta - 4)^2}{(\delta^2 - 9)(\delta^2 - 1)} - 1. \qquad (13)$$

The square root of Eq. 13 is also plotted in Fig. 11.

4.6 Form of the Psychophysical Function

If, as in Theorem 2, we assume that the magnitude scale is (approximately) proportional to the underlying psychophysical function ψ, then Stevens' results clearly suggest that ψ must be a power function for intensive continua. The question facing the theoretician is whether this empirical result can be arrived at from some more primitive considerations. We shall present two theories, neither of which we feel is really satisfactory.

The first is suggested by a study that Plateau (1872) reported in which he gave a pair of painted disks, one black and one white, to each of eight artists and asked them to return to their studios and paint a grey "midway" between the two. The resulting productions were "presque identique" in spite of the fact that they were painted under widely varying conditions. The range of reflectances from the two patches must have been great, yet "midway" was about the same for all eight artists. What had remained fixed, of course, was the ratio of the reflectances from the patches, and so the identical greys suggested that equal stimulus ratios must have induced equal sensation ratios. Given this generalization from one observation, it follows that sensation must be a power function of intensity. The formal statement and proof are the following:

Theorem 4. *If ψ is a positive, real-valued, continuous function of a positive real variable and if for any s, s', t, t' for which $s/t = s'/t'$, it follows that $\psi(s)/\psi(t) = \psi(s')/\psi(t')$; then $\psi(s) = \alpha s^\beta$, where $\alpha > 0$.*

PROOF. The second part of the hypothesis is clearly equivalent to saying that there is a function f such that when $s/t = z$ then $\psi(s)/\psi(t) = f(z)$. Rewriting, if $s = tz$, then $\psi(s) = \psi(tz) = \psi(t)f(z)$. Note that for $t = 1$,

$\psi(z) = \psi(1) f(z)$. If we define $u(s) = \psi(s)/\psi(1)$, then this condition can be restated as

$$\begin{aligned} u(sz) &= \frac{\psi(sz)}{\psi(1)} \\ &= \frac{\psi(s) f(z)}{\psi(1)} \\ &= \frac{\psi(s)\, \psi(z)}{\psi(1)^2} \\ &= u(s)\, u(z). \end{aligned}$$

It is well known that the only continuous solutions to this functional equation are of the form s^β; setting $\alpha = \psi(1) > 0$, we have

$$\begin{aligned} \psi(s) &= \psi(1)\, u(s) \\ &= \alpha s^\beta. \end{aligned}$$

This argument is subject to exactly the same criticisms as Fechner's equal jnd assumption (Sec. 2, Chapter 4); it merely replaces an untested postulate about equal differences by an equally untested one about equal ratios. For this reason we do not believe that it is a satisfactory rationalization of the power function.

A second argument that has sometimes been interpreted as a theory for the psychophysical function is given in Luce (1959). He points out that if (1) the stimulus scale is a ratio scale, (2) the sensation scale is also a ratio scale, (3) the function ψ relating them is single valued and continuous, (4) stimulus values are not multiplied by a dimensional constant in such a way that their product is independent of the unit chosen, and (5) an admissible change of scale for the stimulus variable produces only an admissible change of scale for the sensation variable, then ψ must satisfy the functional equation

$$\psi(ks) = K(k)\, \psi(s),$$

where k represents the unit of s and $K(k)$, the corresponding unit of the sensation scale. From this it is easy to show that ψ must be a power function.

This argument seems unsatisfactory in two respects: it prejudges the question whether sensations form a ratio scale—which, however, is certainly suggested by the data and must be the case if the generalization function depends only upon ratios of sensation values—and, more important, it assumes that the psychophysical function can be stated in terms of the physical scale without bringing in dimensional constants that cancel out the physical units. This is simply not true of many physical laws (e.g., the decay laws), although it is of some (e.g., Ohm's and Newton's laws), and so it seems unwise to invoke it as an a priori assumption here.

In our opinion, therefore, theoretical work on why the psychophysical function seems so often to be the power relation continues to be needed. It must be kept in mind that the reasoning should not be too pervasive because it is not at all clear that the power function is the correct psychophysical function for nonprothetic (metathetic) continua.

4.7 Relations to Other Experiments

If we are correct in supposing that the same fundamental response mechanism underlies all psychophysical experiments, it should be possible to predict aspects of one set of data from any of the others. Many of these connections have not yet been explored, but a few theoretical results can be derived about the connections between recognition and magnitude estimation experiments and some experimental-theoretical results are known about the relation between category and magnitude scales.

Following Stevens, let us suppose that

$$\psi(s) = \alpha(s - \gamma)^{\beta}$$

and, as suggested by Theorem 3, that

$$\zeta(x, y) = \begin{cases} \left(\dfrac{x}{y}\right)^{\delta}, & x \leqslant y \\ \left(\dfrac{x}{y}\right)^{-\delta}, & x > y. \end{cases}$$

Then, by the definition of η,

$$\begin{aligned} \eta(s, t) &= \zeta[\psi(s), \psi(t)] \\ &= \begin{cases} \left[\dfrac{\psi(s)}{\psi(t)}\right]^{\delta}, & s \leqslant t \\ \left[\dfrac{\psi(s)}{\psi(t)}\right]^{-\delta}, & s > t \end{cases} \\ &= \begin{cases} \left(\dfrac{s - \gamma}{t - \gamma}\right)^{\beta\delta}, & s \leqslant t \\ \left(\dfrac{s - \gamma}{t - \gamma}\right)^{-\beta\delta}, & s > t. \end{cases} \end{aligned}$$

Here we have a possible hint why it has proved so difficult to separate the psychophysical function from the generalization function. If both are power functions, as assumed above, then so is their composite, and so no single class of experiments is likely to suggest that two distinct functions are involved.

Assuming a two-stimulus, two-response recognition experiment with no bias, the probability of a correct response is given by

$$p(C) = \frac{1}{1 + \eta(s, t)}.$$

If $s < t$ and if we choose the probability cutoff of π, the equation $p(C) = \pi$ yields

$$\eta(s, t) = \left(\frac{s - \gamma}{t - \gamma}\right)^{\beta\delta} = \frac{1 - \pi}{\pi},$$

and so the recognition π-jnd is given by

$$t - s = (s - \gamma)\left[\left(\frac{\pi}{1 - \pi}\right)^{1/\beta\delta} - 1\right], \tag{14}$$

or in logarithmic (db) measure

$$10 \log_{10} \left(\frac{t - \gamma}{s - \gamma}\right) = \frac{10}{\beta\delta} \log_{10} \left(\frac{\pi}{1 - \pi}\right). \tag{15}$$

To get an idea of the size of the recognition jnd predicted from magnitude estimation data, let us suppose that we are working with 1000-cps tones. From Table 1 we see that β is approximately 0.3. If the standard deviation of the response generalization function lies between 0.2 and 0.4, then we see from the standard deviation curve in Fig. 11 that $5 < \delta < 8$. By taking $\pi = 0.75$ as the usual cutoff and assuming stimuli well above threshold so that we can forget about γ, substitution in Eq. 15 yields a predicted stimulus difference of 2.0 to 3.2 db. Relevant data to check this prediction do not seem to exist.

As Rosner (1961) first pointed out in a closely related context, there may be some difficulties with this argument. We see that it leads to a generalized Weber law (Sec. 1.4, Chapter 4) for the recognition jnd (Eq. 14); however, because γ is always positive for magnitude estimation data, the extra constant in the Weber law is subtracted. This is just opposite to what is needed to fit Weber's law to discrimination data (Sec. 1.4, Chapter 4). No one has reported recognition jnd data, and we cannot be sure that they behave in the same way as discrimination data, but it certainly seems to be a reasonable conjecture. If so, something must be wrong with our argument, at least for very small stimuli. Incidently, use of McGill's correction to the power function (Eq. 7) does not materially alter these remarks.

Concerning the relation between category and magnitude estimation data, it has been well known for some time that the simple mean category scale (Sec. 3.2) is moderately like the logarithm of the corresponding magnitude scale, but there are consistent deviations from a simple logarithmic relation. Recently, Galanter, and Messick (1961) have shown that for the loudness of bursts of noise the Thurstonian category scale based on the equation of categorical judgment and using unequal variances is, to a good approximation, the logarithm of the magnitude scale. Torgerson (1960b) presented similar results for estimations of greyness using the simple mean category scale.

On the basis of his data, Torgerson (1960a) suggested that there is but one psychophysical function underlying both category and magnitude estimation scales—whether you ask the subject to judge differences or ratios, he does the same thing, but depending upon what you ask he does or does not make a logarithmic transformation. This is an interesting hypothesis, but we do not believe that any existing data really prove it. Moreover, our attempts to work out theoretical predictions for the mean category scale from the magnitude estimation model have led to messy equations that are not very revealing.

Assuming that Torgerson's hypothesis is confirmed and that some appropriate category scale is in fact the logarithm of the psychophysical function, as obtained from magnitude estimation data, does this mean that the two methods are equally good? Some seem to feel that it does, but, even if we ignore the instability of category data, we cannot agree. The category scales involve two free parameters, corresponding to a zero and unit, whereas the psychophysical function estimated by the magnitude scale appears to have only one unestimable parameter. The other parameters, the exponent β and "threshold" γ, can be estimated from the data. To be sure, we do not yet understand just what the exponent means or what it is related to, but there can be no doubt that an estimated constant reveals more about a subject than does our arbitrary selection of a zero.

5. DISTANCES

In a number of the response models that we have discussed in the preceding two chapters as well as in this one, parameters arose that were attached to pairs of stimuli rather than to single stimuli. For the simpler stimulus parameters, the scaling problem is, in principle, straightforward: how do the scale values depend upon physical measures of the stimuli? When parameters are associated with pairs of stimuli, matters are somewhat more complicated. There is still nothing like the same understanding of these structures as there is of the simpler scales.

Without exception, we have assumed that either the parameters themselves (in the case of Thurstonian models) or their negative logarithms (in the case of the choice models) behave like measures of distance in the following sense:

Definition 1. *A function* d: $\mathcal{S} \times \mathcal{S} \to$ *real numbers is said to be a* distance measure *if for all* $x, y, z \in \mathcal{S}$,

1. $d(x, y) = d(y, x)$,
2. $d(x, y) \geqslant 0$ and $d(x, y) = 0$ *if and only if* $x = y$,
3. $d(x, z) \leqslant d(x, y) + d(y, z)$.

Two broad classes of questions come to mind. First, is there really any reason to expect measures of distance to arise when subjects make judgments about stimuli? Second, if so, what more can be said about such a measure; for example, can it be treated as the natural distance metric of an Euclidean r-space for some value of r? Given that it can, how can we determine the value of r and the coordinates of the points in the space that correspond to particular stimuli?

The question of a rationalization has been attacked by Restle (1959) following a point of view that is familiar from stimulus sampling theory in learning (Chapter 10, Vol. II). We turn to it first.

5.1 A Rationalization for Distance

Restle supposes that a finite set $\mathcal{A}$ of possible stimulus aspects exists. These aspects can be thought of as a list of the various properties that stimuli may possess and that are relevant to the organism under consideration. Each stimulus x has and is characterized by its set X of aspects ($X \subseteq \mathcal{A}$); by "characterize" we mean that stimuli x and y have the same set of aspects if and only if they are the same stimulus, that is, $X = Y$ if and only if $x = y$. Because the aspects may differentially influence the judgment being made, it is reasonable to suppose that each type of judgment generates its own measure function over the subsets of $\mathcal{A}$.

Definition 2. *A function* m: $2^{\mathcal{A}} \to$ *real numbers, where* $2^{\mathcal{A}}$ *is the set of subsets of* $\mathcal{A}$, *is said to be a* measure *if*

1. *for all* $X \subseteq \mathcal{A}$, $m(X) \geqslant 0$,
2. $m(\emptyset) = 0$, *where* $\emptyset$ *is the empty set*,
3. *for all* X, $Y \subseteq \mathcal{A}$, $m(X \cup Y) = m(X) + m(Y) - m(X \cap Y)$.

Let it be clear that from a formal, axiomatic point of view it does not matter what, if any, intuitions we have about the set of aspects and the measure function over them, but that from the point of view of the psychology assumed it matters a great deal. We feel that this scheme, like the stimulus-sampling theory which it closely resembles, is evasive at the intuitive level. Various assumptions other than Restle's are possible, and they are not clearly inferior to his. Moreover, it seems no more intuitively acceptable to us to assume the existence of sets of aspects and a measure over them than to assume directly the existence of distances between pairs of stimuli, which is what the aspects and measure are intended to justify. Apparently, not everyone feels as we do.

Definition 3. *Let $x, y \in S$ and let $X, Y \subseteq \mathscr{A}$ be their associated aspect sets. If m is a measure over $2^{\mathscr{A}}$, the quantity*[4]

$$d(x, y) = m[(X - Y) \cup (Y - X)] \tag{16}$$

is called the aspect distance *between x and y.*

This appears to be a sensible measure of the dissimilarity between x and y because it is the aspect-measure of the set of aspects in which the two stimuli differ.

Lemma 2

$$\begin{aligned} d(x, y) &= m[(X \cap \bar{Y}) \cup (Y \cap \bar{X})] && (17)\\ &= m(X \cap \bar{Y}) + m(Y \cap \bar{X}) && (18)\\ &= m(X - Y) + m(Y - X) && (19)\\ &= m[(X \cup Y) - (X \cap Y)] && (20)\\ &= m(X) + m(Y) - 2m(X \cap Y). && (21) \end{aligned}$$

PROOF. Equation 17 is equivalent to Eq. 16 by the definition of difference.

Equation 18 is equivalent to Eq. 17 by applying property 3 of a measure, noting that the measure of the intersection term is 0 because $(X \cap \bar{Y}) \cap (Y \cap \bar{X}) = (X \cap \bar{X}) \cap (Y \cap \bar{Y}) = \emptyset$ and then using property 2.

Equation 19 is equivalent to Eq. 18 by the definition of difference.

Equation 20 is equivalent to Eq. 17 because $(X \cup Y) - (X \cap Y) = (X \cap \bar{Y}) \cup (Y \cap \bar{X})$ by simple set transformations.

Equation 21 is equivalent to Eq. 20 because $X \cap Y \subseteq X \cup Y$, and so $m[(X \cup Y) - (X \cap Y)] = m(X \cup Y) - m(X \cap Y) = m(X) + m(Y) - 2m(X \cap Y)$.

Theorem 5. *The aspect distance is a distance measure.*

PROOF. By Eq. 21,

$$\begin{aligned} &d(x, y) + d(y, z)\\ &\quad = m(X) + m(Y) - 2m(X \cap Y) + m(Y) + m(Z) - 2m(Y \cap Z)\\ &\quad = m(X) + m(Z) - 2m(X \cap Z) + 2[m(X \cap Z)\\ &\qquad\qquad + m(Y) - m(X \cap Y) - m(Y \cap Z)]\\ &\quad = d(x, z) + 2[m(X \cap Z) + m(Y) - m(X \cap Y) - m(Y \cap Z)]\\ &\quad \geqslant d(x, z) \end{aligned}$$

[4] $X - Y = X \cap \bar{Y}$ denotes the set theoretic difference of X and Y.

provided that $m(X \cap Z) + m(Y) - m(X \cap Y) - m(Y \cap Z) \geqslant 0$. To show this, define the following pairwise disjoint sets (see Fig. 12):

$$\begin{aligned} A &= X \cap \bar{Y} \cap Z \\ B &= X \cap Y \cap Z \\ C &= X \cap Y \cap \bar{Z} \\ D &= \bar{X} \cap Y \cap Z. \end{aligned}$$

Clearly,

$$\begin{aligned} X \cap Z &= A \cup B \\ X \cap Y &= B \cup C \\ Y \cap Z &= B \cup D \\ Y &\supseteq B \cup C \cup D. \end{aligned}$$

Thus,

$$\begin{aligned} & m(X \cap Z) + m(Y) - m(X \cap Y) - m(Y \cap Z) \\ & \qquad \geqslant m(A) + m(B) + m(B) + m(C) + m(D) \\ & \qquad\qquad\qquad - m(B) - m(C) - m(B) - m(D) \\ & \qquad \geqslant 0, \end{aligned}$$

as was to be shown.

Observe from this last proof that

$$d(x, y) + d(y, z) = d(x, z) \tag{22}$$

if and only if

$$m(A) = 0 \quad \text{and} \quad m[Y - (B \cup C \cup D)] = 0. \tag{23}$$

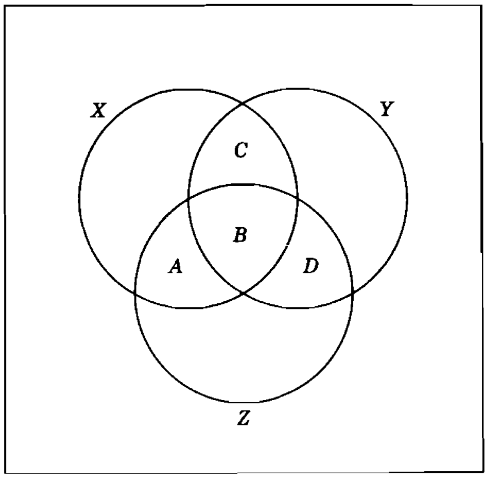

Fig. 12. A graphic representation of the eight pairwise disjoint sets discussed in Theorem 5.

If we suppose, with no psychological loss of generality, that for any $U \subseteq \mathscr{A}$, $m(U) = 0$ implies $U = \emptyset$, then Eq. 23 implies

$$A = \emptyset \quad \text{and} \quad Y = B \cup C \cup D.$$

It is not difficult to see that this is equivalent to

$$X \cap Z \subseteq Y \subseteq X \cup Z,$$

which in a certain reasonable sense can be interpreted as meaning that Y is between X and Z. This leads to the following definition:

Definition 4. *For stimuli x, y, and z which are all different, y is said to be* between *x and z, written $x\,|y|\,z$, if and only if*

$$X \cap Z \subseteq Y \subseteq X \cup Z. \tag{24}$$

We have already proved the following theorem.

Theorem 6. *If $x\,|y|\,z$, then Eq.* 22 *holds; the converse is true if and only if $U = \emptyset$ whenever $m(U) = 0$.*

Corollary. *If $x\,|y|\,z$, then $d(x, z) > d(x, y)$ and $d(x, z) > d(y, z)$.*

PROOF. Equation 22 and property 1 of a distance measure.

Care must be taken not to overrespond to Theorem 6, which seems to suggest that betweenness acts in the same way here as it does on an ordinary one-dimensional mathematical continuum. Because Eq. 22 holds only for triples of stimuli, we cannot conclude anything about larger sets of stimuli until we have shown it to hold. Unfortunately, not everything we would like to be true is true. For example, a usual form of extrapolation is the following:

Conjecture: *If $w\,|x|\,y$ and $x\,|y|\,z$, then $w\,|x|\,z$ and $w\,|y|\,z$.*

Counter Example. Let $W = \{1, 2\}$, $X = \{2, 3\}$, $Y = \{3, 4\}$, $Z = \{4, 5\}$, then $w\,|x|\,y$ because $W \cap Y = \emptyset \subseteq X \subseteq W \cup Y$ and $x\,|y|\,z$ because $X \cap Z = \emptyset \subseteq Y \subseteq X \cup Z$. But not $w\,|x|\,z$ because $X = \{2, 3\} \nsubseteq \{1, 2, 4, 5\} = W \cup Z$ and not $w\,|y|\,z$ because $Y = \{3, 4\} \nsubseteq \{1, 2, 4, 5\} = W \cup Z$.

We next show that a form of interpolation is true.

Theorem 7. *If $w\,|x|\,z$ and $x\,|y|\,z$,* then *$w\,|y|\,z$.*

PROOF. Because $w\,|x|\,z$, $W \cap Z \subseteq X$ and because $x\,|y|\,z$, $X \cap Z \subseteq Y$, so $W \cap Z = W \cap Z \cap Z \subseteq X \cap Z \subseteq Y$. Because $x\,|y|\,z$, $Y \subseteq X \cup Z$ and because $w\,|x|\,z$, $X \subseteq W \cup Z$, so $Y \subseteq X \cup Z \subseteq W \cup Z \cup Z = W \cup Z$. Thus, by definition, $w\,|y|\,z$.

Another result of interest is this one:

Theorem 8. *If $x\,|y|\,z$, then $z\,|y|\,x$ but not $y\,|x|\,z$ or any of the other permutations of the three symbols.*

PROOF. $z\,|y|\,x$ follows immediately from the definition of $x\,|y|\,z$ and the commutativity of union and intersection.

Suppose $y\,|x|\,z$ as well as $x\,|y|\,z$, then we know that

$$X \cap Z \subseteq Y \subseteq X \cup Z$$
$$Y \cap Z \subseteq X \subseteq Y \cup Z.$$

From the first we have

$$X \cap Z = X \cap Z \cap Z \subseteq Y \cap Z$$

and from the second

$$Y \cap Z = Y \cap Z \cap Z \subseteq X \cap Z,$$

so $X \cap Z = Y \cap Z$. In like manner,

$$Y \cap \bar{Z} \subseteq (X \cup Z) \cap \bar{Z} = (X \cap \bar{Z}) \cup (Z \cap \bar{Z}) = X \cap \bar{Z}$$

and

$$X \cap \bar{Z} \subseteq (Y \cup Z) \cap \bar{Z} = (Y \cap \bar{Z}) \cup (Z \cap \bar{Z}) = Y \cap \bar{Z},$$

so $X \cap \bar{Z} = Y \cap \bar{Z}$. Thus

$$X = (X \cap Z) \cup (X \cap \bar{Z}) = (Y \cap Z) \cup (Y \cap \bar{Z}) = Y,$$

that is, $x = y$, contrary to the definition of $x\,|y|\,z$.

Although, in the general case, the betweenness relation is not strong enough to patch together sets of ordered stimuli, there are special assumptions about the aspect sets for which it is possible. The simplest case is a set of stimuli $x_1, x_2, \ldots, x_n$ such that $X_1 \subset X_2 \subset \ldots \subset X_n$. This is known as a *monotone sequence of sets*, and it obviously has the property that if $i < j < k$ then $x_i\,|x_j|\,x_k$. The aspect set of a stimulus higher than another in the series is obtained from the lower one by adding new aspects. This appears to correspond to the definition of stimuli on what Stevens and Galanter (1957) have called *prothetic continua*, for in their terms stimulation is added to stimulation to give rise to a growth in sensation. Typical prothetic continua are those that are sometimes called intensive: sound intensity which gives rise to loudness, light intensity which gives rise to brightness, etc. In the model, aspect distance is additive when the aspect sets form a monotone sequence of sets (Theorem 6).

A somewhat more general notion, which includes monotone sequences of sets as a special case, is a sequence of sets generated in the following way from any three mutually disjoint sets of aspects, A, B, and C. The first stimulus in the array has the aspect set $A \cup B$. The second is obtained by removing some elements from B, but not from A, and adding some from C. The third is obtained by removing more from B, but not from A or those added from C, and by adding more from C, and so on. The ith aspect set is of the form $A \cup B_i \cup C_i$, where $B_i \subseteq B$ and $C_i \subseteq C$. The formal definition can be given as follows:

Definition 5. *A sequence of distinct sets $X_1, X_2, \ldots, X_n$ form a* linear array of sets *if there exist sets $A, B_1, \ldots, B_n, C_1, \ldots, C_n$, such that*

1. $A \cap B_1 = A \cap C_n = B_1 \cap C_n = \emptyset$,
2. for $i < j$, $B_j \subseteq B_i$, and $C_i \subseteq C_j$,
3. $X_i = A \cup B_i \cup C_i$.

In terms of the betweenness notion, the following seems to capture what we might mean by a linear array of stimuli.

Definition 6. *A sequence of distinct stimuli $x_1, x_2, \ldots, x_n$ form a* linear array of stimuli *if for all i, j, k such that $i < j < k$, then $x_i \,|x_j|\, x_k$.*

Theorem 9. *A sequence of stimuli form a linear array of stimuli if and only if their aspect sets form a linear array of sets.*

PROOF. Suppose, first, that the aspect sets form a linear array, and consider $i < j < k$. Because $B_k \subseteq B_j \subseteq B_i$ and $C_i \subseteq C_j \subseteq C_k$, we have

$$\begin{aligned} X_i \cap X_k &= (A \cup B_i \cup C_i) \cap (A \cup B_k \cup C_k) \\ &= A \cup B_k \cup C_i \\ &\subseteq A \cup B_j \cup C_j \\ &= X_j \end{aligned}$$

and

$$\begin{aligned} X_i \cup X_k &= (A \cup B_i \cup C_i) \cup (A \cup B_k \cup C_k) \\ &= A \cup B_i \cup C_k \\ &\supseteq A \cup B_j \cup C_j \\ &= X_j. \end{aligned}$$

Thus, by Def. 4, $x_i \,|x_j|\, x_k$.

Now, suppose that $x_1, x_2, \ldots, x_n$ form a linear array of stimuli. Define

$$A = X_1 \cap X_n, \quad B_i = X_i \cap X_1 \cap \bar{A}, \quad \text{and} \quad C_i = X_i \cap X_n \cap \bar{A}.$$

Observe that

$$B_1 = X_1 \cap \bar{A} = X_1 - A \quad \text{and} \quad C_n = X_n \cap \bar{A} = X_n - A,$$

and so

$$B_i = X_i \cap B_1 \quad \text{and} \quad C_i = X_i \cap C_n.$$

We show that these sets satisfy the conditions of Def. 5.

1. $A \cap B_1 = A \cap X_1 \cap \bar{A} = \emptyset$,
   $A \cap C_n = A \cap X_n \cap \bar{A} = \emptyset$,
   $B_1 \cap C_n = (X_1 \cap \bar{A}) \cap (X_n \cap \bar{A}) = X_1 \cap X_n \cap \bar{A} = A \cap \bar{A} = \emptyset.$

2. Suppose $i < j$. For $i = 1$, $B_1 \supseteq X_j \cap B_1 = B_j$. For $i > 1$,

$$\begin{aligned} B_j &= X_1 \cap \bar{A} \cap X_j \\ &= (X_1 \cap \bar{A}) \cap (X_1 \cap X_j) \\ &\subseteq X_1 \cap \bar{A} \cap X_i \\ &= B_i, \end{aligned}$$

because, for $1 < i < j$, $x_1 |x_i| x_j$, which in turn implies $X_1 \cap X_j \subseteq X_i$. A similar argument shows that $C_i \subseteq C_j$.

3. Consider

$$\begin{aligned} A \cup B_1 \cup C_n &= A \cup (X_1 - A) \cup (X_n - A) \\ &= X_1 \cup X_n \\ &\supseteq X_i, \end{aligned}$$

because, for $1 < i < n$, $x_1 |x_i| x_n$. Thus

$$\begin{aligned} X_i &= X_i \cap (A \cup B_1 \cup C_n) \\ &= (X_i \cap A) \cup (X_i \cap B_1) \cup (X_i \cap C_n) \\ &= A \cup B_i \cup C_i, \end{aligned}$$

because $X_i \cap A \supseteq X_1 \cap X_n \cap X_1 \cap X_n = X_1 \cap X_n = A$.

Corollary. *If $x_1, x_2, \ldots, x_n$ form a linear array of stimuli, then for $i < j < k$, $d(x_i, x_k) = d(x_i, x_j) + d(x_j, x_k)$.*

PROOF. $X_i - X_k = (B_i - B_k)$ and $X_k - X_i = (C_k - C_i)$ by Def. 5, and so

$$d(x_i, x_k) = m(B_i - B_k) + m(C_k - C_i).$$

Because $B_i \supseteq B_j \supseteq B_k$ and $C_i \subseteq C_j \subseteq C_k$,

$$\begin{aligned} d(x_i, x_k) &= m(B_i - B_j) + m(B_j - B_k) + m(C_k - C_j) + m(C_j - C_i) \\ &= d(x_i, x_j) + d(x_j, x_k). \end{aligned}$$

It is evident that a linear array of sets is a monotone sequence of sets if $B_1 = \emptyset$ and that any monotone sequence is a linear array.

The structure of a linear array of sets involves the substitution of some aspects for others to get from one stimulus to another, which seems to correspond to the characterization given by Stevens and Galanter (1957) of a metathetic continuum. Examples are pitch, hue, etc. One problem may exist in making these identifications between the empirically defined scales and those of the aspect model. If the model is correct, the class of metathetic continua includes the prothetic as a special case, or, put another way, the dividing line between the two classes of continua need not be sharp, although in nature it may be. Certain borderline arrays of sets simply may not have counterparts among the psychological continua.

In summary, then, Restle has shown that there is a way to assign distances to pairs of stimuli provided that one assumes that a measure over aspect sets exists and that Eq. 16 defines the distance. Because we have no experimental identification either of aspects or aspect measures, it is anyone's guess whether this notion of distance has any relation to those that have arisen in the response models.

5.2 The Embedding Problem

Assuming that we have stimulus parameters that satisfy the properties of a distance measure (Def. 1), the next question is whether they can be interpreted as distances in some familiar space. If so, that space may then be taken as a multidimensional representation of the stimuli, and, presumably, one would then attempt to discover the relations between coordinates of the space and physical attributes of the stimuli. Little has been done on this last problem.

Although several authors (Attneave, 1950; Galanter, 1956) have suggested that non-Euclidean spaces may be appropriate, little research has been reported on anything other than Euclidean embeddings. There are dangers in limiting ourselves to this familiar space. Because of sampling errors, it is never possible to demand that the estimated distances rigidly meet the mathematical criteria for a particular embedding; and because the statistical features have not been fully worked out, a good deal of judgment is involved in deciding whether a particular embedding is appropriate. But because our judgments are likely to be influenced by our presystematic intuitions about the nature of the space and the arrangement of the stimuli in it, there is some fear that we are simply perpetuating the errors of naïve Euclidean intuition.

The main theorems describing the conditions under which error-free distances can be embedded in an r-dimensional Euclidean vector space were first stated and proved by Young and Householder (1938).[5] Consider a set of n points a_i lying in an Euclidean vector space, the origin of which coincides with say, the nth point. Let α_i be the vector from the nth to the ith point and let α_{ij} be the component of α_i along the jth coordinate. The matrix $A = [\alpha_{ij}]$ has rank r equal to the dimensionality of the space spanned by the given points, which is also the rank of $B = AA'$. It is easy to see that the elements of B, b_{ij}, are the dot product of α_i with α_j, and so by elementary properties of vectors

$$b_{ij} = \tfrac{1}{2}[d^2(i, n) + d^2(j, n) - d^2(i, j)] = d(i, n)\, d(j, n) \cos \theta_{ijn}, \qquad (25)$$

[5] To follow their arguments, it is necessary that the reader be familiar with certain basic ideas and results from matrix theory.

where the d's are distances between points. Thus we have proved the following theorem:

Theorem 10. *The dimensionality of a set of n points in an Euclidean vector space with distances $d(i,j)$ is equal to the rank of the $n - 1$ square matrix B whose elements are defined by Eq. 25.*

By Eq. 25, it is clear that B is symmetric, which with $B = AA'$ implies B is positive semidefinite. Conversely, suppose B is positive semidefinite; then we know that it has only nonnegative latent roots. Hence, by a well-known theorem, there exists an orthogonal matrix Q such that

$$B = QL^2Q' = (QL)(QL)',$$

where L is a diagonal matrix of the latent roots of B. If we set $A = QL$, then we have the coordinates of the vectors of the embedding, which proves the following theorem.

Theorem 11. *A necessary and sufficient condition that a set of points with distances $d(i,j) = d(j,i)$ be embeddable in an Euclidean vector space is that the matrix B whose elements are defined by Eq. 25 be positive semidefinite; the embedding is unique up to translations and rotations.*

The condition of positive semidefiniteness implies that the determinant of each of the 2×2 principal minors must be positive, which in turn is equivalent to the triangle inequality (Part 3 of Def. 1). The remaining requirements are, in essence, generalizations of this property.

For error-free data, it does not matter which stimulus is selected as the origin. To be sure, the matrix B depends upon this choice, but the ranks of all the B matrices are equal and the embedding is the same except for translations and rotations. With real data, however, matters are not so simple. Each choice yields a slightly different embedding. Torgerson (1952, 1958) suggested a procedure of locating the origin at the centroid of the several points and finding a single "average" B^* matrix, which can then be factored by the methods of factor analysis to find the matrix of components A^*. A discussion and description of some empirical uses of these methods with distances obtained by the similarity model of Sec. 1.2 is given by Torgerson (1958).

6. CONCLUSIONS

Whereas we tend to be theory-rich and data-poor in discrimination research, the reverse seems to be true in scaling. To be sure, when a close formal analogy exists between a scaling method and an identification experiment (e.g., between similarity scaling and discrimination), scaling theories are often easily constructed simply by reinterpreting the theory for the corresponding identification experiment; but when the analogies are not close, either the theories are not satisfactory, as in category scaling, or they are not well developed, as in magnitude estimation.

To some extent, the directness by which the scaling procedures yield scales has led some psychologists to the view that little in the way of theory is really needed; the methods seem to get at what one wants without any fancy theoretical indirection. In our view, however, these methods presuppose certain theoretical results that need to be explicitly stated and studied. Specifically, until adequate models are evolved, the following three classes of questions, which seem basic in all of psychophysics, are not likely to receive anything like final answers.

1. If we confine our attention to a single physical variable, such as sound energy, and to a single relevant judgment, such as loudness, just how many distinct sensory mechanisms are needed to account for the experimentally observed behavior? The models that we have studied suggest that at least two are needed. For example, in the choice model of Sec. 4.4 the two mechanisms are represented mathematically by the psychophysical scale and by the generalization function. The question is whether these two are sufficient to explain the results from, for example, recognition, discrimination, similarity, category, bisection, and magnitude estimation experiments or whether more mechanisms are needed. We shall probably not answer this question soon because of the complex way in which these functions combine with the ubiquitous, but poorly understood, biasing function to predict the subject's responses. In fact, we suspect that the answer will come only with the rather complete confirmation of an elaborate response theory. If so, then the problem of sensory measurement will have proved to be more analogous to, say, electrical measurement, which was perfected only as electrical theory itself became well understood, rather than to the measurement of length and weight, which was relatively well developed long before any adequate physical theories involving these quantities were stated.

2. In all of the theories in which stimulus functions (or parameters) are defined over pairs of stimuli—in the choice models for recognition, similarity, and magnitude estimation experiments and in the Thurstonian model for similarity experiments—what mathematical structure do they exhibit? Uniformly, we have assumed that they or a simple transformation of them behave like distances in an Euclidean space, but these assumptions have not been carefully tested. They need to be because psychological similarity simply may not be a distance notion.

3. In all psychophysical data there is considerable evidence that subjects bias their responses and that these biases can be affected by presentation

probabilities, payoffs, and other experimentally manipulable factors. In the theories that we have discussed in this and the two preceding chapters these biases are represented as a function defined over responses. The nature of this function—its dependence on things that we can manipulate experimentally—is not known. For some purposes this does not seem to matter critically, although it has some inherent interest to many people, but for other purposes, such as ascertaining the relation between the magnitude estimation scale and the underlying psychophysical function, knowledge of at least the general mathematical form of the biasing function seems to be essential if we are to avoid being misled about sensory scales.

References

Aczél, J. On mean values. *Bulletin of the American Mathematical Society*, 1948, **54**, 392–400.

Adams, E. W., & Messick, S. An axiomatic formulation and generalization of successive intervals scaling. *Psychometrika*, 1958, **23**, 355–368.

Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516–556.

Fernberger, S. W. On absolute and relative judgments in lifted weight experiments. *Amer. J. Psychol.*, 1931, **43**, 560–578.

Galanter, E. An axiomatic and experimental study of sensory order and measure. *Psychol. Rev.*, 1956, **63**, 16–28.

Galanter, E., & Messick, S. The relation between category and magnitude scales of loudness. *Psychol. Rev.*, 1961, **68**, 363–372.

Garner, W. R. Advantages of the discriminability criteria for a loudness scale. *J. acoust. Soc. Amer.*, 1958, **30**, 1005–1012.

Helm, C., Messick, S., & Tucker, L. *Psychophysical law and scaling models.* Princeton, N.J.: Educational Testing Service Research Bulletin, 59–1, 1959.

Herrnstein, R. J., & van Sommers, P. Method for sensory scaling with animals. *Science*, 1962, **135**, 40–41.

Luce, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81–95.

Luce, R. D. A choice theory analysis of similarity judgments. *Psychometrika*, 1961, **26**, 151–163.

McGill, W. J. The slope of the loudness function: A puzzle. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: theory and application.* New York: Wiley, 1960. Pp. 67–81.

Messick, S., & Abelson, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–17.

Michels, W. C., & Doser, Beatrice T. Rating scale method for comparative loudness measurements. *J. acoust. Soc. Amer.*, 1955, **27**, 1173–1180.

Parducci, A. Incidental learning of stimulus frequencies in the establishment of judgment scales. *J. exp. Psychol.*, 1956, **52**, 112–118.

Pfanzagl, J. *Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens.* Schrift. d. Stat. Inst. d. Univ. Wien, Neue Folge Nr. 1, 1959.(a)

Pfanzagl, J. A general theory of measurement: applications to utility. *Naval Research Logistics Quarterly*, 1959, **6**, 283–294.(b)

Plateau, M. H. Sur la mesure des sensations physique, et sur la loi qui lie l'intensité de ces sensations a l'intensité de la cause excitante. *Bull. acad. roy. Belg.*, 1872, **33,** 376–388.

Restle, F. A metric and an ordering on sets. *Psychometrika*, 1959, **24**, 207–220.

Rosner, B. S. Psychophysics and neurophysiology. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 4. New York: McGraw-Hill, 1961. Pp. 280–333.

Saffir, M. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2,** 179–198.

Steinberg, J. C., & Gardner, M. B. The dependence of hearing impairment on sound intensity. *J. acoust. Soc. Amer.*, 1937, **9,** 11–23.

Stevens, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56,** 246–250.

Stevens, S. S. *Handbook of Experimental Psychology*. New York: Wiley, 1951.

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

Stevens. S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **54,** 177–196.(a)

Stevens, S. S. Some similarities between hearing and seeing. *Laryngoscope*, 1958, **68,** 508–527.(b)

Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57,** 201–209.

Stevens, S. S. The psychophysics of sensory function. *Amer. Sc.*, 1960, **48,** 226–253.

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

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory Communication*. New York: Wiley, 1961. Pp. 1–33.(b)

Stevens, S. S., & Galanter, E. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54,** 377–411.

Titchener, E. H. *Experimental Psychology*. Vol. II, Part II (Instructor's Manual). New York: Macmillan, 1905.

Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17,** 401–409.

Torgerson, W. S. A law of categorical judgment. In L. H. Clark (Ed.), *Consumer Behavior*. Washington Square: New York Univ. Press, 1954. Pp. 92–93.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Torgerson, W. S. *Distances and ratios in psychophysical scaling*. Massachusetts Institute of Technology, Lincoln Laboratory Report, 48G–0014, 1960.(a)

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen and S. Messick (Eds.), *Psychological scaling*. New York: Wiley, 1960.(b)

Wever, E. G., & Zener, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35,** 466–493.

Young, G., & Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, **3,** 331–333.