# 4

# Discrimination [1,2,3]

R. Duncan Luce

*University of Pennsylvania*

Eugene Galanter

*University of Washington*

# Contents

# Discrimination

"To note the differences between. To set apart as different; differentiate; distinguish. To observe a difference; distinguish. To make a distinction; treat unequally or unfairly." These are the ordinary meanings of *discriminate*. Traditionally, in psychophysics, it has had a special meaning which we take up first; later, in Sec. 5, we consider experiments in which the responses are more nearly coordinate with the dictionary meaning.

In many early discrimination experiments the stimuli differed on only one physical dimension, and the subject judged not only whether the two stimuli of a presentation differed but which he believed to be the larger, or more intense, etc. Thus his possible responses were "larger," "same," and "smaller," or something equivalent. Later studies deviated even more from the usual meaning of discriminate in that the "same" response was omitted, for evidence accumulated that subjects were quite unstable in their definition of that category (Thompson, 1920; Boring, 1920). When the only permitted responses are "larger" and "smaller," or the like, we call it a *forced-choice* procedure—"forced" because the subject is not permitted to say that two stimuli are the same, even when they are or seem to be. The models we discuss first are designed to account for forced-choice data.

The exact responses used, words or other signals, depend upon how the experiment is implemented, but there must be just two when pairs of stimuli are presented and they must be unambiguously related to the physical characteristics of the stimuli. For example, if one stimulus is presented after the other, then the subject can be asked to designate whether the first or the second is the larger or, equivalently, whether the second is larger or smaller than the first.[4] If the stimuli, for example, color patches, are presented simultaneously, then spatial designations such as "up" and "down" or "right" and "left" are used or the patches are labeled 1 and 2 or $A$ and $B$, etc. In any event, a label that is unambiguous to both the subject and the experimenter is used.

In terms of the notation described in Chapter 2, we have a set $\mathscr{S}$ composed of stimuli that are assumed to differ on only one continuous physical dimension. This dimension can be anything we choose—mass, energy, frequency, etc.—as long as there is a physical measure that defines a weak

[4] The logical equivalence of these responses does not imply that two experiments which differed only in this way would yield identical data. See pp. 194, 224 ff.

ordering $\geqslant$ over $\mathscr{S}$, that is, $\geqslant$ must be a connected, reflexive, and transitive binary relation. In the following we shall say that $\mathscr{a}$ is "larger than" $\mathscr{a}'$ if $\mathscr{a} > \mathscr{a}'$ under the given physical ordering. In the simplest case the presentation set $S$ is a subset of $\mathscr{S} \times \mathscr{S}$, and the abstract response set $R$ is simply $\{1, 2\}$ (or any other pair of symbols). If we let

$$I_1 = \{\langle \mathscr{a}^1, \mathscr{a}^2 \rangle \mid \mathscr{a}^1, \mathscr{a}^2 \in \mathscr{S} \text{ and } \mathscr{a}^1 > \mathscr{a}^2\}$$

$$I_2 = \{\langle \mathscr{a}^1, \mathscr{a}^2 \rangle \mid \mathscr{a}^1, \mathscr{a}^2 \in \mathscr{S} \text{ and } \mathscr{a}^2 > \mathscr{a}^1\},$$

then the identification function is

$$\iota(1) = I_1$$
$$\iota(2) = I_2.$$

This, as we pointed out in Chapter 2, is an example of a partial identification experiment.

The most inclusive two-alternative, forced-choice design is called *pair comparisons*, and it is characterized by $S = \mathscr{S} \times \mathscr{S}$. If $\mathscr{S}$ consists of $k$ stimuli, a total of $k^2$ ordered pairs can be formed. For example, when $\mathscr{S} = \{\mathscr{a}, \mathscr{a}', \mathscr{a}''\}$, then there are the nine ordered pairs: $\langle \mathscr{a}, \mathscr{a} \rangle$, $\langle \mathscr{a}, \mathscr{a}' \rangle$, $\langle \mathscr{a}, \mathscr{a}'' \rangle$, $\langle \mathscr{a}', \mathscr{a} \rangle$, $\langle \mathscr{a}', \mathscr{a}' \rangle$, $\langle \mathscr{a}', \mathscr{a}'' \rangle$, $\langle \mathscr{a}'', \mathscr{a} \rangle$, $\langle \mathscr{a}'', \mathscr{a}' \rangle$, $\langle \mathscr{a}'', \mathscr{a}'' \rangle$. To say that the first stimulus is larger than the second when $\langle \mathscr{a}, \mathscr{a}' \rangle$ is presented seems to be equivalent to saying that the second is larger than the first when $\langle \mathscr{a}', \mathscr{a} \rangle$ is presented, and so one might be tempted to present only one of the two pairs. It is not, however, empirically true that the results are the same: the order of presentation affects the behavior. Just how it affects behavior is a question we study later; at present we need only realize that it does.

For an interesting number of stimuli, say 10 to 100, there are quite a few ordered pairs, 100 to 10,000. What makes these numbers unpleasant is that it is insufficient to present each pair just once, for subjects do not always respond the same way each time a pair is presented. To get reasonably stable relative frequencies of these choices, we must present each pair many times—"many" being of the order of 100 to 1000. Thus a modest pair comparison experiment entails 10,000 observations and a big one, 10 million. The need to reduce these numbers, at least the second one, is evident.

The clue to doing so lies in the fact that inconsistencies exist only when two stimuli are physically not very different—just what we mean by "not very different" will become clear later. There is surely little point, then, in presenting pairs hundreds of times when the response is certain. This suggests that we pick one or more stimuli, which are called *standards*, and study only pairs formed from other stimuli, called *comparison* stimuli, in the near neighborhood of each standard. A standard stimulus has,

therefore, its own set of comparison stimuli. Usually, though not always, the standard is presented first, and a comparison, second, and the subject is asked to judge whether the second is greater or less than the first. Although it is not necessary, in most studies an odd number of comparison stimuli are used, with the physically middle one often being identical to the standard. This type of design is known as the *method of constant stimuli,* although a more appropriate name seems to be the "method of standard stimuli."

## 1. RESPONSE MEASURES AS FUNCTIONS OF STIMULUS MEASURES

### 1.1 The Psychometric Function

So that subjects are not free to indulge their taste for consistency—which most seem to have—we are careful in most constant stimuli experiments not to group all like presentations together. Some irregular presentation schedule, often a simple random one, is used. When it comes to data analysis, however, we want to deal with the responses to each pair as a unit, and so from now on our problem is to account for these sets of responses to presentation pairs.

Changing notation slightly, let $x$ denote the numerical value of the relevant physical dimension of a typical comparison stimulus and $s$, the numerical value of the standard. Suppose, to be specific, that each $\langle s, x \rangle$ pair is responded to 100 times. It will surprise no one that, when $x$ is sufficiently smaller (on whatever physical variable differentiates them) than $s$, then all or almost all of the subject's responses are that it is smaller. He is equally correct when $x$ is sufficiently larger than $s$. As we move up the series of comparison stimuli from the smallest to the largest, we find that his reports first decrease in consistency, reach a peak at which roughly half his responses are that the comparison stimulus is larger, and then become increasingly more consistent until he always says that it is larger. What is interesting is the patterning to the inconsistency: the proportion of "larger" responses changes rather smoothly from 0 to 1.

To see this, we may plot the proportion as a function of the physical measure of the stimulus. Although we could plot the physical measure as the independent variable, it is more usual to use either the dimensionless linear scale $(x - s)/s$ or the dimensionless logarithmic scale $\log (x/s) = \log x - \log s$. Both have the virtue of setting the zero at $s$, which is sensible because we are studying the neighborhood of $s$. A typical data plot is shown in Fig. 1.
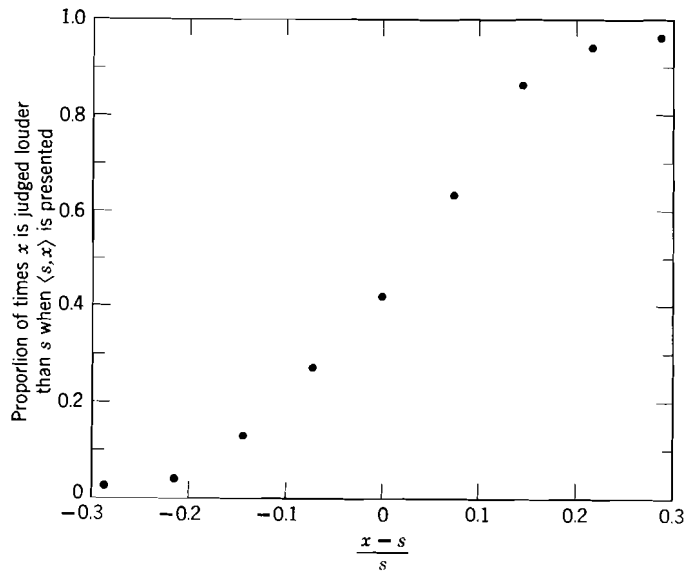
Fig. 1. Loudness discrimination data (E. Galanter, unpublished). Each data point is based upon 105 observations collected over three sessions. The standard stimulus was at 50 db *re* 0.0002 dynes/cm$^2$.

The $S$ (or sigmoidal or ogival) shape of these points is typical of discrimination data, as are the slight irregularities. Were we to introduce another comparison stimulus between two actually used, we anticipate that the new proportion would usually fall between the two old data points, approximately on a smooth S-curve faired to the original data points. Of course, we cannot be sure exactly where it will fall. In two different runs of 100 observations we do not generally get exactly the same proportions, but we would be surprised if they were very different from each other or from the faired curve. And data confirm such conjectures.

These considerations lead us to postulate that for each value $x$, $x > 0$ of the comparison stimulus and for each $s$, $s \geqslant s_0 > 0$ there exists a probability $p(2 \mid \langle s, x \rangle)$ governing the response that $x$ is judged larger than $s$. Clearly, $p(1 \mid \langle s, x \rangle) = 1 - p(2 \mid \langle s, x \rangle)$. The number $s_0$ is interpreted as a thresholdlike quantity, somewhat larger than the usual absolute or detection threshold. Moreover, we postulate that these probabilities satisfy:

Assumption 1. *Strict Monotonicity:* *if* $x < y$, *then* $p(2 \mid \langle s, x \rangle) \leqslant p(2 \mid \langle s, y \rangle)$, *and,* *when* $p(2 \mid \langle s, x \rangle)$ *and* $p(2 \mid \langle s, y \rangle) \neq 0$ *or* 1, *then* $p(2 \mid \langle s, x \rangle) < p(2 \mid \langle s, y \rangle)$.

Assumption 2.  *Differentiability:* $p(2 \mid \langle s, x \rangle)$ *is a differentiable and, therefore, continuous function of each of its arguments* $x$ *and* $s$ *except, possibly, at a finite number of points.*

Assumption 3.  *Limiting behavior:*

$$\lim_{x \to \infty} p(2 \mid \langle s, x \rangle) = 1 \quad and \quad \lim_{x \to 0} p(2 \mid \langle s, x \rangle) = 0.$$

Such a function, assuming that it exists, is called a *psychometric function* (Urban, 1907).

The usual and unbiased estimate of $p(2 \mid \langle s, x \rangle)$ is the proportion of times that the subject responded $x$ is larger when the pair $\langle s, x \rangle$ was presented.

The reader should be fully aware that these three assumptions about the psychometric function are not easily tested.  The inherent binomial variability of the data means that the observed proportions can have the opposite order from the actual probabilities and that they can be rather widely separated when the probabilities are actually quite close.  A numerical example is revealing.  We know that when $N$ independent observations are made, each with probability $p$ of success, then the expected proportion of successes is $p$ and the standard deviation of the proportion is $\sqrt{p(1-p)/N}$.  If $p = 0.50$ and $N = 100$, the standard deviation is 0.05.  If $p' = 0.55$, the standard deviation is nearly the same.  Therefore, we should not be unduly surprised to observe proportions of 0.56 and 0.53, respectively, which is an apparent violation of monotonicity, or of 0.43 and 0.59, which might suggest a violation of continuity.  Only by very careful experimentation and statistical analysis can one detect such violations, unless they are quite gross, which they are not.  In the absence of contrary evidence, experimental or theoretical, monotonicity and continuity seem to be sensible assumptions.

## 1.2  The PSE, CE, and jnd

Let $p_y(2 \mid \langle s, y \rangle)$ be the partial derivative of $p(2 \mid \langle s, y \rangle)$ with respect to $y$, which by Assumption 2 exists.  Thus

$$p(2 \mid \langle s, x \rangle) = \int_0^x p_y(2 \mid \langle s, y \rangle) \, dy,$$

where the constant of integration is 0 by Assumption 3.  The monotonicity assumption implies

$$p_x(2 \mid \langle s, x \rangle) \geqslant 0,$$

and Assumption 3 implies

$$\int_0^\infty p_y(2 \mid \langle s, y \rangle) \, dy = \lim_{x \to \infty} p(2 \mid \langle s, x \rangle) = 1.$$
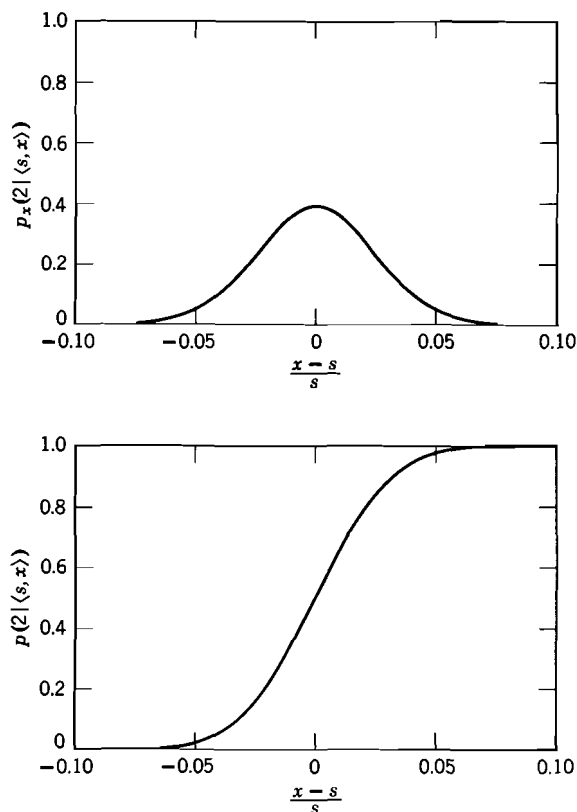
Fig. 2. Typical plots of $p_x(2 \mid \langle s, x \rangle)$ versus $(x - s)/s$ and of its cumulative $p(2 \mid \langle s, x \rangle)$ versus $(x - s)/s$. The density function is normal and its cumulative is, in this context of a psychometric function, often called a normal ogive.

Thus $p_y$ is a probability density function and $p$ is its distribution function. Moreover, $p$ is an S-shaped function if and only if $p_y$ is unimodal. An example of this relation is shown in Fig. 2.

Two important features of a density function, especially a unimodal one, are a measure of central tendency, such as the mean or median, and a measure of dispersion, such as the standard deviation or the interquartile range. In discrimination studies it has been customary to use the median and interquartile range rather than the mean and standard deviation because the former pair makes weaker assumptions about the metric properties of the independent variable (Urban 1907, Boring 1917).

The median, $x_{1/2}$, of the distribution $p_y$ is by definition that point such that

half the area under $p_y$ lies to the left of it. Because the total area is 1, this means that $x_{1/2}$ is characterized by

$$p(2 \mid \langle s, x_{1/2} \rangle) = \int_0^{x_{1/2}} p_y(2 \mid \langle s, y \rangle) \, dy = \tfrac{1}{2}. \tag{1}$$

In the discrimination context, $x_{1/2}$ is often called the *point of subjective equality* and is abbreviated PSE. As we noted earlier, order of presentation matters, so in general $x_{1/2} \neq s$. The difference, $x_{1/2} - s$, is called the *constant error*, or CE, which is in a way a misnomer because it can be altered by certain experimental manipulations.

The interquartile range is the interval $x_{3/4} - x_{1/4}$, where $x_{3/4}$ is the stimulus value above $x_{1/2}$ such that $\tfrac{1}{4}$ of the area under $p_y$ lies between these two points and $x_{1/4}$ is the corresponding point below $x_{1/2}$. More succinctly, they are defined by

$$p(2 \mid \langle s, x_{3/4} \rangle) = \tfrac{3}{4} \quad \text{and} \quad p(2 \mid \langle s, x_{1/4} \rangle) = \tfrac{1}{4}.$$

Most often, psychophysicists work with half the interquartile range, that is,

$$\text{jnd} = \frac{x_{3/4} - x_{3/4}}{2}, \tag{2}$$

which is called the *just noticeable difference* or the *difference limen* (DL). We use the first term.[5] The jnd is really an algebraic approximation to a probabilistic structure, as was first emphasized by Urban. We go into its structure as an algebraic entity in Sec. 3.3.

Graphically, the several quantities just defined are shown in Fig. 3.

It should be realized that the cutoffs of $\tfrac{3}{4}$ and $\tfrac{1}{4}$ in the definition of the jnd are arbitrary; other values, $\pi$ and $1 - \pi$, where $\tfrac{1}{2} < \pi < 1$, could have been used, and we would speak of the $\pi$-jnd at $s$. When the value of $\pi$ is not specified, it is taken for granted that it is $\tfrac{3}{4}$.

## 1.3 Estimating the PSE and jnd

To estimate the PSE and the jnd when the mathematical form of the psychometric function is known, it seems appropriate first to find that function of the given class which gives the "best" fit to the data and then to calculate these quantities according to these definitions. Various definitions of "best," such as least squares and maximum likelihood, can

[5] Classically, the jnd and the DL were distinguished as a theoretical term and a statistic of data, respectively. However they are now often used interchangeably, and so we adopt the first term to refer to the statistic because of its mnemonic value.
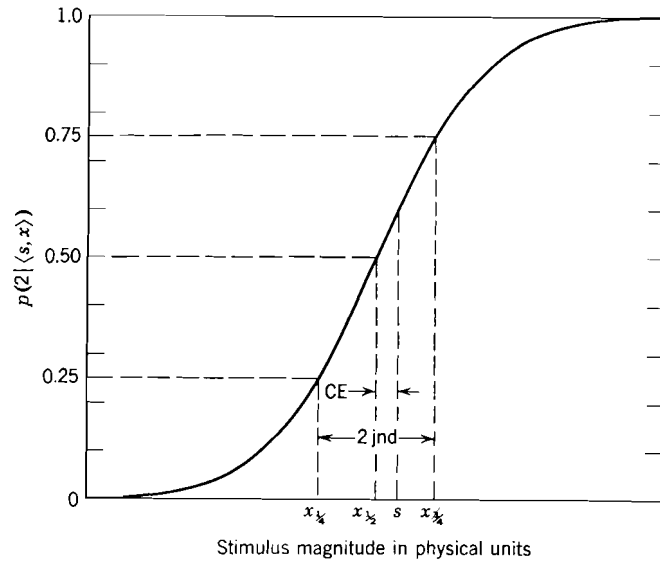
Fig. 3. A typical psychometric function [$p(2 \mid \langle s, x \rangle)$ versus $x$] showing the CE and the jnd.

be used, and solutions are known for certain classes of functions (see Chapter 8). In practice, however, these calculations are usually tedious, and a rather simpler procedure can be used, which, although not optimal, seems quite satisfactory for many purposes.

The data uniformly suggest that the psychometric function is very nearly a straight line in a region approximately one jnd, or a little more, above and below the PSE. So, instead of using a more precise psychometric function, we may calculate a best-fitting straight line to those data points between about 0.2 and 0.8; usually a least squares procedure is used, although with practice one can become quite skillful doing it graphically by eye. If the equation of this line is $p = ax + b$, then it is easy to see that

$$\hat{x}_{1/2} = \frac{1 - 2b}{2a}$$

and that

$$\widehat{jnd} = \frac{1}{4a}.$$

Thus the jnd and the slope are inversely related: good discrimination means a small jnd and a large slope; poor discrimination, a large jnd and a small slope.

## 1.4 Study of the jnd: Weber functions and Weber's law

The psychometric function tells us something about local behavior, namely, about the nature of discrimination in the neighborhood of the standard $s$. One can ask: how does this function vary with experimental manipulation, with, among other things, changes in the standard? Because it is difficult to describe how an entire function varies, the question is usually reduced to asking how the CE and the jnd vary with experimental conditions; however, somewhat different questions are posed in the two cases. In this section we take up the jnd and postpone discussing the CE until Sec. 4.

Of two techniques to determine the jnd, the one giving the smaller value is considered the better, for the primary interest has been the absolute limits of discrimination. For this reason, modifications of equipment, instructions, outcomes, or procedures are deemed desirable and are incorporated into the experiment if they reduce the size of the jnd, but for the most part their effects, as such, upon the size of the jnd are not studied. This does not mean that such questions are meaningless or even uninteresting but only that other problems have seemed more important. This leaves us only one major manipulation for studying the jnd, the value of the standard stimulus.

If we estimate the jnd at $s$, $\widehat{jnd}$ $(s)$, for several different values of $s$, we may then construct a plot such as that shown in Fig. 4. Actually, it is not easy to find reports of jnds estimated at different values of $s$ using the method of constant stimuli. For example, the data of Fig. 4 (Miller, 1947) for the intensity of random noise were obtained by the quantal method described in Sec. 6.3 of Chapter 3, with the jnd being defined as that increment in intensity which is detected 50 per cent of the time. The curves shown in Fig. 6 below were obtained by the method of limits in which the variable stimulus is systematically increased and decreased until a difference is found that is noted 50 per cent of the time. Most experimentalists assume that, aside from a constant factor, the same function would be obtained by the method of constant stimuli or by one of the other, somewhat more convenient methods, but this assumption has yet to be adequately checked. For present purposes, we will assume that it is correct.

The regularity of the data points in Fig. 4 suggests that a continuous, monotonically increasing function underlies them. This function we denote by jnd $(s)$ and call it the jnd-function (of the stimulus variable).

Such functions were first studied by E. H. Weber in the second quarter
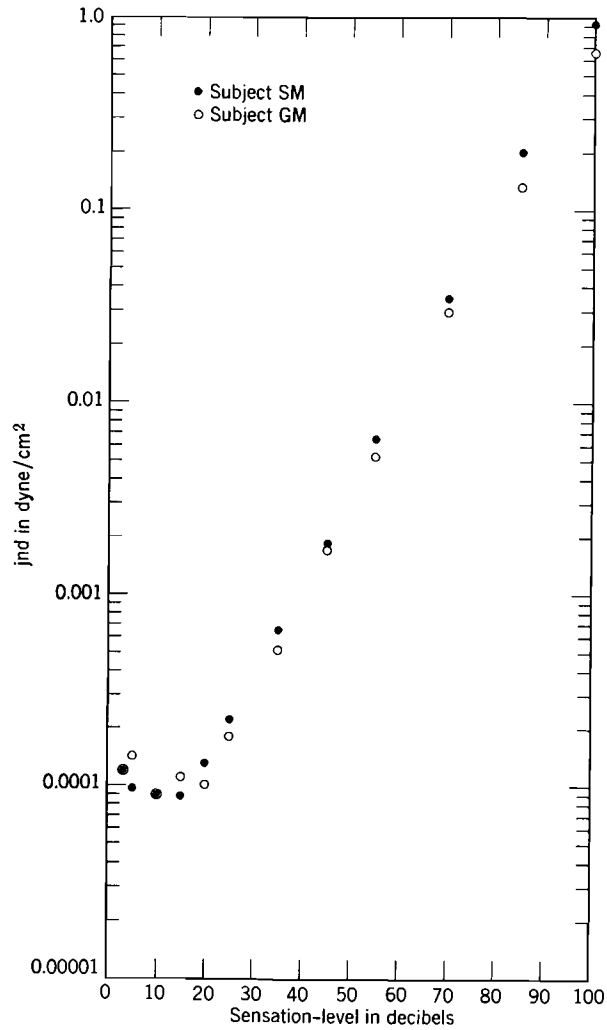
Fig. 4. Intensity discrimination of random noise for two subjects using the quantal method. The jnd is measured in dynes/cm² and is plotted on a logarithmic scale. The stimulus is measured in so-called sensation level units which is decibels*re* subject's threshold (which is approximately 10 db *re* 0.0002 dynes/cm²). Adapted by permission from Miller (1947, p. 612).

of the nineteenth century for cutaneous sensitivity discrimination. On the basis of his data, Weber suggested that for intensity variables, at least, the jnd-function is of the form

$$\text{jnd } (s) = Ks, \tag{3}$$

where the constant $K > 0$ is called the *Weber fraction* and the proposed empirical relation is called *Weber's law*. Usually one plots the data as $\widehat{\text{jnd}}\ (s)/s$ versus $s$ or log $s$, in which case the points lie on a horizontal line $K$ units above the $x$-axis when Weber's law is correct. Actually, most data show an initial drop, then a relatively flat region, and sometimes a final rise for very "large" stimuli (Cobb, 1932, Holway & Pratt, 1936). The final rise has been debated, and some believe that it can be accounted for by difficulties in the experimental procedure or equipment, but there is little doubt about the initial dip.

To cope with the dip, Fechner proposed and G. A. Miller (1947) later revived a *generalized Weber's law* of the form

$$\text{jnd}\ (s) = Ks + C, \tag{4}$$

where $K > 0$. In Fig. 5 Miller's jnd-data for intensity discrimination of white noise are shown along with a fitted curve of the form of Eq. 4. The approximation is good.

Guilford (1932) proposed as a substitute for Eq. 3 or 4,

$$\text{jnd}\ (s) = Ks^n,$$
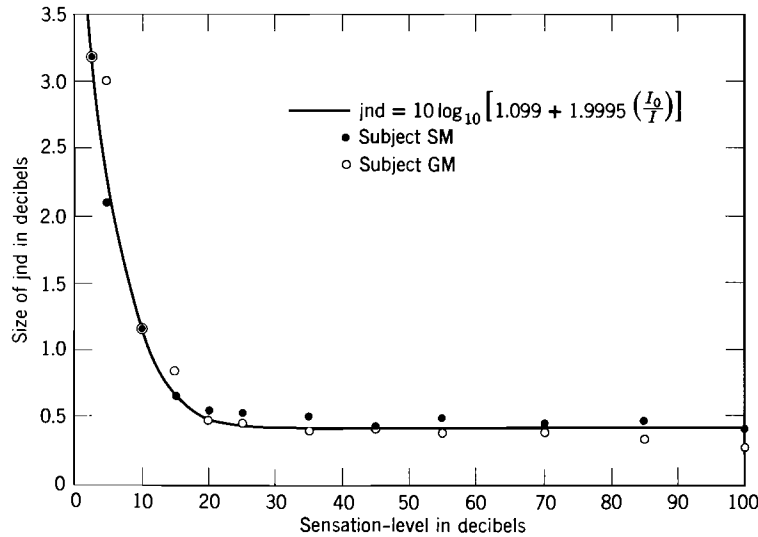


Fig. 5. The data of Fig. 4 are replotted with the jnd measured in decibels, i.e., if at intensity $I$ the jnd measured in dynes/cm² is $\Delta(I)$, then the jnd in db $= 10 \log_{10} \left[ 1 + \dfrac{\Delta(I)}{I} \right]$.

The theoretical curve is the generalized Weber law with the numerical constants shown. Adapted by permission from Miller (1947, p. 612).
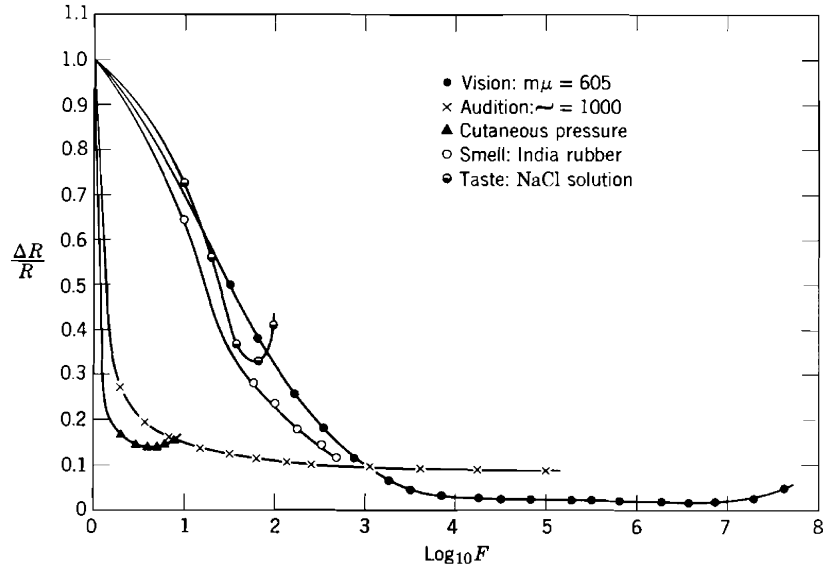
Fig. 6. Weber fractions versus sensation-level of the stimulus (in logarithmic units) for five senses obtained by the method of limits. The Weber fractions have been normalized to be unity at threshold. Adapted by permission from Holway & Pratt (1936, p. 337).



Fig. 7. A typical psychometric function showing the values of the Weber functions $\Delta(s, 1 - \pi)$ and $\Delta(s, \pi)$ at $s$ as well as the CE and the $\pi$-jnd.

where $n > 0$. Hovland (1938) presented data for which this is a somewhat better approximation than Eq. 4. Householder and Young (1940) point out, however, that the law that holds depends upon the physical measure of the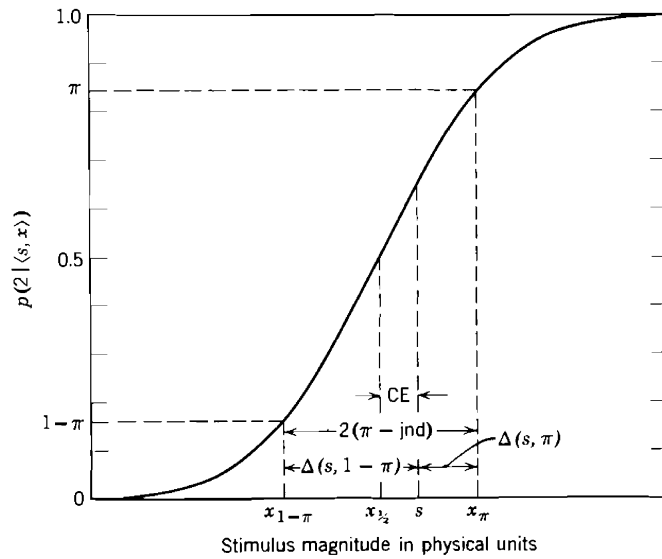 stimulus used. Under very general mathematical conditions it is possible to find a measure for which Weber's law is correct for a single subject, but such a transformation will work for all subjects only under very special circumstances.

Table 1    Values of the Weber Fraction

|  | Weber Fraction |
|---|---|
| Deep pressure, from skin and subcutaneous tissue, at about 400 grams | 0.013 |
| Visual brightness, at about 1000 photons | 0.016 |
| Lifted weights, at about 300 grams | 0.019 |
| Tone, for 1000 cps, at about 100 db above absolute threshold | 0.088 |
| Smell, for rubber, at about 200 olfacties | 0.104 |
| Cutaneous pressure, on an isolated spot, at about 5 grams/mm | 0.136 |
| Taste, for saline solution at about 3 moles/liter concentration | 0.200 |

Taken from Boring, Langfeld, & Weld (1948) p. 268.

It hardly need be mentioned that such jnd-functions are empirically defined only from the absolute threshold up to either the upper threshold, as for pitch, or just short of intensity levels that damage the receptors, as for sound and light intensity.

Table 1 presents estimates of the Weber fraction $K$ for several continua, and Fig. 6 presents plots of several Weber functions.

For the theoretical work to follow, it is convenient to use a somewhat different measure of dispersion than the jnd. Let $\pi$ be a number such that $0 < \pi < 1$. We define the function $\Delta(s, \pi)$ by the property

$$p[2 \mid \langle s, s + \Delta(s, \pi) \rangle] = \pi,$$

that is,

$$\Delta(s, \pi) = x_\pi - s. \tag{5}$$

Depending upon the value of $\pi$, $\Delta(s, \pi)$ may be positive or negative. For a fixed $\pi$, we shall call $\Delta(s, \pi)$ a *Weber function* of $s$.

It should be noted that $CE = \Delta(s, \frac{1}{2})$ but that for other values of $\pi$, $\Delta(s, \pi)$ is a measure of dispersion. In fact, if the $CE = 0$, then jnd $(s) = \frac{1}{2}[\Delta(s, \frac{3}{4}) - \Delta(s, \frac{1}{4})]$, and in general $\pi$-jnd $(s) = \frac{1}{2}[\Delta(s, \pi) - \Delta(s, 1 - \pi)]$. (See Fig. 7.)

No one has paid much heed to the fact that the empirically calculated

jnd and the Weber functions used in theoretical work are defined slightly differently. When, for example, the jnd is found to satisfy the generalized Weber law, one usually takes for granted that the Weber function does too, that is,

$$\Delta(s, \pi) = K(\pi)s + C(\pi). \tag{6}$$

But, if this is true, then

$$CE = \Delta(s, \tfrac{1}{2}) = K(\tfrac{1}{2})s + C(\tfrac{1}{2}), \tag{7}$$

which is more than is usually intended. Possibly this CE relation is correct—we know of no relevant data—but, if it is not, then, because $\Delta(s, \pi)$ — CE plays a closer role to the jnd for symmetric psychometric functions than does $\Delta(s, \pi)$, we should assume Weber's law for the former and not for the latter.

## 2. FECHNERIAN SCALING[6]

### 2.1 The Problem of Scaling

With the publication of *Elemente der Psychophysik* in 1860, G. T. Fechner not only founded psychophysics as a science but introduced a theoretical idea which has dominated the field in the intervening years. It is this. As the magnitude of stimulation is varied, for example, as the sound energy of a tone is varied, we appreciate the change as a closely parallel subjective change, in this case in what is called loudness. We say that stimulation produces a sensation peculiar to it, and most of us allege that sensations vary continuously and monotonically with the usual physical magnitudes involved. They do not, however, seem to vary linearly: a 10-unit change of intensity at low levels does not seem to be the same size as a 10-unit change at high intensities.

Fechner, and many after him, wanted to know exactly how "sensation intensity" varies with physical magnitude. The problem is not one of straight-forward experimentation because the question is not an empirical one until we know how to define and measure sensation. The question of definition, and therefore of measurement, is, we should judge, still unsettled and probably will not be finally resolved until psychophysics is a more nearly perfected chapter of science than it is today; nonetheless, we shall explore a number of different answers that have been proposed and seriously entertained. It may be well to pause a moment here to sketch our general view of the matter, to make our biases known, before we discuss the views of others.

[6] To many, the problems of this section are of no more than historical interest. But they certainly are that, and so we feel it worthwhile to devote space to their careful analysis. If one wishes, the section can be omitted with little loss of continuity.

First, it is not clear what behavioral observations constitute a measure of sensation. This role has been claimed for one or another class of observations, but none has had such clear face validity that it has not been vigorously questioned by some careful students of the field. Second, although it seems reasonable to suppose that physiological measures of sensation may one day be found, little is known about them today. Third, in several mathematical theories of discrimination (and other choice behavior) it is not particularly difficult to introduce numerical scales that, on the one hand, allow us to reconstruct the behavioral data rather efficiently and, on the other hand, are monotonically, but not usually linearly, related to physical intensity. Fourth, although each of these scales is a possible candidate for the unknown measure of sensation, no clear criterion for choice yet exists. It is probably unwise for scientific purposes (engineering applications are another matter) to attempt to make any terminological decision until more is known. Aside from a question of terms, it is by no means certain which, if any, of these scales is truly useful. If a scale serves no purpose other than as a compact summary of the data from which it was calculated, if it fails to predict different data or to relate apparently unrelated results, that is, if it is not an efficient theoretical device, then it is worth but little attention and surely we should not let it appropriate such a prized word as "sensation." If, however, a scale is ever shown to have a rich theoretical and predictive role, then the scientific community can afford to risk the loss of a good word. At present, no scale meets this criterion.

The scales of this section are the most traditional in psychology, and at the same time the most unlikely. What we shall be doing is to take highly local information, as given by the psychometric function, and from it attempt to generate scales over the whole range of the physical variable. In recent years a school of thought, led by S. S. Stevens, has condemned this processing of confusion (or noise or variability) as, on the face of it, irrelevant to measures of magnitude. We take up his alternative method, magnitude estimation, more fully in Chapter 5; here it suffices to describe Fechner's idea. Earlier critiques of Fechner's whole scheme of psychophysics can be found in Cobb (1932), James (1890), and Johnson (1929, 1930, 1945).

## 2.2 Fechner's Problem

Let a particular value $\pi$ be chosen and suppose that $\Delta(s, \pi)$ is the resulting Weber function. Fechner posed the question: for what (reasonably smooth) transformations of the physical scale is the transformed
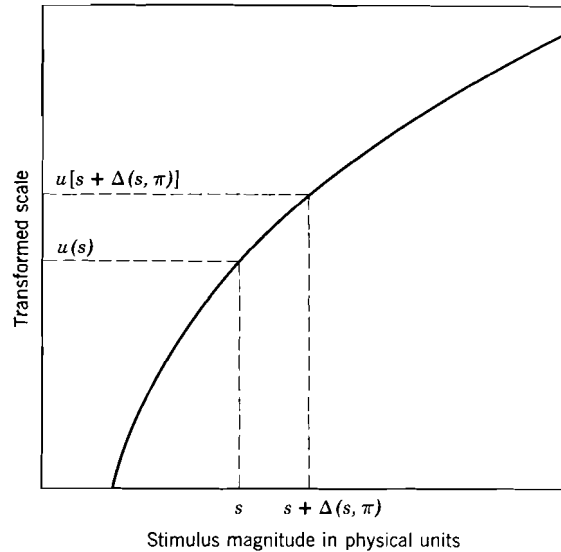
Fig. 8. A typical monotonic transformation showing the interval on the transformed scale corresponding to the value $\Delta(s, \pi)$ of the Weber function at $s$ on the original scale.

Weber function a constant independent of $s$? This defining property of the "sensation" scale is sometimes loosely phrased by saying that at the level of sensations jnds are all equal. Mathematically, if $u(s)$ is a transformation of the physical scale, then in terms of the $u$-variable the Weber function is simply $u[s + \Delta(s, \pi)] - u(s)$. This is easily seen in Fig. 8.

**Fechner's Problem.** *For a fixed $\pi$, $0 < \pi < 1$, find those "smooth" strictly monotonic functions $u$ such that for all $s > s_0$*

$$u[s + \Delta(s, \pi)] - u(s) = g(\pi), \tag{8}$$

*where $g$ is a strictly monotonic increasing function of $\pi$ and is independent of $s$.*

Two comments. First, we have explicitly included the vague word "smooth," its exact definition to be decided upon later when we see how the problem develops. Second, the assumption that $g$ is a strictly monotonic function follows from the assumption that the psychometric function is strictly monotonic for $0 < p < 1$.

Is this characterization of the unknown scale $u$ a self-evident truth, an empirical assumption, or a definition? Fechner, we believe, believed it to be the first or possibly the second. The era of self-evident truths having largely passed, some today regard it as an assumption, and the remainder,

a definition (Boring, 1950; Cobb, 1932). Without an independent measure of $u$, it is difficult for us to see how it can be considered an assumption. But no matter, let us accept the problem as it stands without worrying about its philosophic status.

The task Fechner faced was to solve the functional equation (8). The method he attempted to use—see Boring (1950) or Luce & Edwards (1958)—involved certain approximations which allow Eq. 8 to be transformed into a first-order linear differential equation. Such differential equations are well known to have solutions which are unique except for an additive constant of integration provided that certain weak assumptions are met. However, because $g(\pi)$ is unspecified, the unit of $u$ is also free, so by this argument $u$ is an interval scale (i.e., it is unique up to positive linear transformations). Certainly, such reasoning would be acceptable had the problem been initially cast as the differential equation, but it was not, and for other classes of functional equations one must investigate carefully both the questions of the existence and uniqueness of solutions. Neither can ever be taken for granted with unfamiliar equations.

Because $\pi$ is a fixed quantity in Fechner's problem, we drop it from the notation. The following result solves the uniqueness question.

Theorem I.   *Let $u^*$ be a strictly monotonic solution to Eq. 8; then $u$ is another solution to Eq. 8 if and only if*

$$u(s) = u^*(s) + F[u^*(s)],$$

*where $F$ is a periodic function of period $g$, that is, $F(x + g) = F(x)$ for all $x$.*

PROOF.   First, suppose $F$ is periodic with period $g$; then by the definition of $u$,

$$u[s + \Delta(s)] - u(s) = u^*[s + \Delta(s)] + F\{u^*[s + \Delta(s)]\} - u^*(s) - F[u^*(s)].$$

Because $u^*$ is a solution to Eq. 8, we know that

$$u^*[s + \Delta(s)] = g + u^*(s),$$

and because $F$ is periodic with period $g$

$$F\{u^*[s + \Delta(s)]\} = F[u^*(s) + g]$$
$$= F[u^*(s)].$$

Thus

$$u[s + \Delta(s)] - u(s) = u^*[s + \Delta(s)] + F[u^*(s)] - u^*(s) - F[u^*(s)]$$
$$= g,$$

as was to be shown.

Now, suppose $u$ and $u^*$ are both solutions to Eq. 8, and let $w = u - u^*$. Then

$$w[s + \Delta(s)] - w(s) = u[s + \Delta(s)] - u^*[s + \Delta(s)] - u(s) + u^*(s)$$
$$= g - g$$
$$= 0.$$

Because $u^*$ is strictly monotonic, its inverse $u^{*-1}$ exists, so we may write

$$s + \Delta(s) = u^{*-1}[u^*(s) + g]$$

by Eq. 8 and

$$s = u^{*-1}[u^*(s)]$$

by the definition of an inverse. Define $F = w(u^{*-1})$. Substituting, we have

$$0 = w[s + \Delta(s)] - w(s)$$
$$= w\{u^{*-1}[u^*(s) + g]\} - w\{u^{*-1}[u^*(s)]\}$$
$$= F[u^*(s) + g] - F[u^*(s)].$$

Thus $F$ is periodic with period $g$, and

$$u = u^* + w$$
$$= u^* + w[u^{*-1}(u^*)]$$
$$= u^* + F(u^*),$$

as was to be shown.

There is trouble. The class of solutions to Eq. 8 is, according to Theorem 1, much too heterogeneous to be acceptable as a scale. If, for example, $u^*$ is a nice smooth function, then $u^* + F(u^*)$ may be quite a lumpy function, such as that shown in Fig. 9. It is difficult to consider these two functions as merely two different representations of the same "subjective scale."

We need a recasting of the problem that permits us to narrow down the set of solutions to, say, an interval scale. This is not difficult to find. Our first phrasing of the problem involved an arbitrary cutoff $\pi$. Now, either the value of the cutoff is completely inessential or the problem holds no scientific interest. If we get one scale for one cutoff, another for a different one, and so on, we can hardly believe that some one of these scales has any inherent importance, whereas the others have none. So we rephrase the problem.

The Revised Fechner Problem. *For every $\pi$, $0 < \pi < 1$, find those "smooth" strictly monotonic functions $u$ that are independent of $\pi$ such that for all $s \geqslant s_0$*

$$u[s + \Delta(s, \pi)] - u(s) = g(\pi), \tag{9}$$

*where $g$ is a strictly monotonic increasing function of $\pi$ and is independent of $s$.*
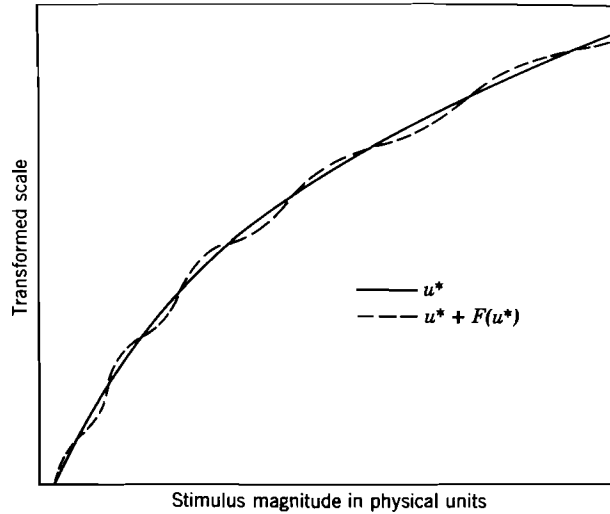
Fig. 9. Examples of solutions to Fechner's problem that are related by a periodic function.

**Theorem 2.** *If $u$ and $u^*$ are two continuous strictly monotonic solutions to Eq. 9, then there exists a constant $b$ such that $u = u^* + b$.*
PROOF. By the preceding theorem, we know that

$$u = u^* + F(u^*),$$

where $F$ must be periodic with period $g(\pi)$ for all $\pi$, $0 < \pi < 1$. Because $g$ is a strictly monotonic function, $F$ has a nondenumerable set of periods which together with its continuity (implied by that of $u$ and $u^*$) means that $F$ must be a constant, as asserted.

Because the unit of $u^*$ is arbitrary, Theorem 2 implies that the solutions to the revised Fechner problem constitute an interval scale.

Before turning to some discussion of the existence of solutions, we show that the revised Fechner problem is identical to a condition that has gone under the lengthy title "equally often noticed differences are equal, unless always or never noticed."

**The Equally-Often-Noticed-Difference Problem.** *Let $p(2 \mid \langle s, x \rangle)$ be a psychometric function which, for $0 < p < 1$, is strictly monotonic in the $x$-argument. Find those strictly monotonic functions $h$ and $u$, if any, such that for all $x > 0$ and $s \geqslant s_0$ for which $p(2 \mid \langle s, x \rangle) \neq 0$ or $1$*

$$p(2 \mid \langle s, x \rangle) = h[u(x) - u(s)].$$ 
(10)

**Theorem 3.** *The equally-often-noticed-difference problem has a solution if and only if the revised Fechner problem has a solution and both problems have the same set of scales.*

PROOF. Suppose the Fechner problem has the solution $u$. For $x$ and $s$ such that $p(2 \mid \langle s, x \rangle) = \pi \neq 1$ or $0$, we know that $x = s + \Delta(s, \pi)$ by the definition of the Weber function. Because $u$ solves the Fechner problem, $u(x) - u(s) = g(\pi)$. Because $g$ is strictly monotonic, it has an inverse $h$, hence

$$h[u(x) - u(s)] = h[g(\pi)]$$
$$= \pi$$
$$= p(2 \mid \langle s, x \rangle),$$

so $u$ solves the equally-often-noticed-difference problem.

Conversely, suppose $u$ solves the equally-often-noticed-difference problem. Let $\pi \neq 0$ or $1$ be given; then by definition of $\Delta(s, \pi)$,

$$p[2 \mid \langle s, s + \Delta(s, \pi) \rangle] = \pi = h\{u[s + \Delta(s, \pi)] - u(s)\}.$$

Because $h$ is strictly monotonic, its inverse, $g$, exists, and so

$$g(\pi) = u[s + \Delta(s, \pi)] - u(s),$$

as was to be shown.

### 2.3 Fechner's Law

It is by no means obvious under what conditions the revised Fechner problem has a solution and, when it has, what its mathematical form is. A variety of sufficient conditions is known, two of which are described in later sections and one here. Moreover, a general expression, in terms of a limit (but not an integral, as Fechner thought) is known for the solution when it exists (see Koenigs, 1884, 1885, or Luce & Edwards, 1958). We shall not present it here.

If the generalized Weber law, $\Delta(s, \pi) = K(\pi)s + C(\pi)$, is empirically correct, then, of course, we are interested in solutions only for that case; fortunately, it is quite well understood.

**Theorem 4.** *If the generalized Weber law holds, a solution exists to the revised Fechner problem provided that $C(\pi)/K(\pi)$ is a constant, $\gamma$, independent of $\pi$; the solution is*

$$u(s) = A \log (s + \gamma) + B,$$

*when $A > 0$ and $B$ are constants.*

PROOF.   $u[s + \Delta(s, \pi)] - u(s) = u[s + K(\pi)(s + \gamma)] - u(s)$

$\qquad\qquad = A \log [s + K(\pi)(s + \gamma) + \gamma] + B - A \log (s + \gamma) - B$

$\qquad\qquad = A \log [1 + K(\pi)].$

Because $u$ is strictly monotonic, and $\Delta(s, \pi) = K(\pi)(s + \gamma)$, it follows easily that $A > 0$.

Observe that this restriction $C(\pi)/K(\pi) = \gamma$ is strong: it means that $\Delta(s, \pi) = K(\pi)(s + \gamma)$, which is tantamount to Weber's original law with the origin of the physical variable shifted to $-\gamma$.

The logarithmic scale is sometimes called *Fechner's law*, although it is not a law in any usual sense of the term—the scale $u$ is not independently defined. Frequently, the set of ideas that we have discussed in this section are grouped together as the *Weber-Fechner problem*, but, as has been repeatedly emphasized in the literature, Weber's and Fechner's laws are independent of one another (e.g., Cobb, 1932).

As we turn to later work, we should keep in mind that Fechner attempted to establish a scale by demanding that subjective jnds on that scale all be equal. In doing this, he made no attempt to specify anything about the form of the psychometric function. When, because of uniqueness problems, we revised his problem to get rid of its dependence upon an arbitrary probability cutoff, it became apparent that solutions exist only when the psychometric functions satisfy rather stringent conditions. It is not surprising, therefore, that we devote considerable attention to the form of these functions.


## 3. UNBIASED RESPONSE MODELS

Partly because of historical interest and partly because it is easier to begin this way, we first take up models that apply only when the order of presentation of the stimuli does not matter. Specifically, we assume that for all $x, y \in \mathscr{S}$, such that $\langle x, y \rangle, \langle y, x \rangle \in s$,

$$p(1 \mid \langle x, y \rangle) = p(2 \mid \langle y, x \rangle). \tag{11}$$

Because in both cases the response designates a particular stimulus, namely $x$, as larger, it is possible and convenient to speak as if the subject had chosen $x$ from the unordered set $\{x, y\}$ and to write $p(x \mid \{x, y\})$ for the common value of Eq. 11. Usually this notation is simplified further by letting $p(x, y) = p(x \mid \{x, y\})$.

## 3.1 The Equation of Comparative Judgment

In 1927 L. L. Thurstone, in his papers "Psychophysical analysis" and "A law of comparative judgment,"[7] took up Fechner's approach from a new point of view, one which has since dominated theoretical work in psychophysics and, even more, psychometrics. Unlike Fechner, Thurstone was concerned with why there is any confusion at all between two stimuli or, to put it another way, why the psychometric function is not a simple jump function from 0 to 1. This surely is what we would expect if repeated presentations of a stimulus always produce the same internal effect and if the same decision rule is always applied to the internal information generated by the two stimuli of a presentation. The facts being otherwise, these cannot both be correct assumptions. A varying decision rule, although a real possibility, is most distasteful because it seems so unmanageable; hence the other assumption is the first to be abandoned.

The problem, then, is to develop a model for the information to which the simple decision rule is applied. Thurstone postulated that the "internal effect" of each stimulus can be summarized by a number, but not necessarily the same number each time the stimulus is presented. Although we need not say why this should happen, plausible reasons exist. Undoubtedly there are small uncontrolled variations in the experimental conditions as well as some internal to the subject himself which affect the appearance of the stimulus. Put in statistical language, the presentation of stimulus $x$ is assumed to result in a random variable $X$ whose range is the real numbers. Thurstone called these random variables *discriminal processes;* their distribution he called the *discriminal dispersion* in (1927b), but in (1927a) he used this term for the standard deviation of the distribution and that has come to be the accepted usage.

It should be noticed that this is exactly the same postulate made by Tanner and his colleagues when analyzing detection and recognition problems (see Chapter 3, Sec. 1.1 for a more complete discussion of this idea).

Thurstone then assumed the following decision rule:

$$Stimulus\ x\ is\ judged \begin{Bmatrix} larger \\ smaller \end{Bmatrix} than\ y\ if\ X \begin{Bmatrix} > \\ < \end{Bmatrix} Y. \qquad (12)$$

[7] These papers, along with many others of Thurstone's, have recently been reprinted in Thurstone (1959). To some extent, Thurstone's ideas stem from those of Müller (1879), Solomons (1900), and Urban (1907). For other expositions of Thurstone's work, see Gulliksen (1946) and Torgerson (1958).

Thus the probability that $x$ is judged larger than $y$ is the same as the probability that $X - Y > 0$, that is,

$$p(x, y) = \text{Pr} (X - Y > 0).\tag{13}$$

To calculate this, we need to know the distribution of the differences $X - Y$, which quantities Thurstone referred to as *discriminal differences*.

To get at this distribution, we consider the single random variable $X$. We suggested that the range of $X$ is not a unique number because of a variety of minor uncontrolled disturbances. If these perturbations are independent and random, and if their number is large, then their over-all sum will be approximately normally distributed. So we assume that $X$ is normally distributed with mean $u(x)$, which Thurstone called the *modal discriminal process*, and standard deviation $\sigma(x)$. Similarly, $Y$ is normally distributed with mean $u(y)$ and standard deviation $\sigma(y)$. Because $x$ and $y$ are usually presented in close spatial and temporal contiguity, it is quite possible that the random variables $X$ and $Y$ are correlated;[8] let the correlation coefficient be $r(x, y)$. It is a well known statistical result that under these conditions $X - Y$ is normally distributed with mean $u(x) - u(y)$ and variance $\sigma^2(x, y) = \sigma^2(x) + \sigma^2(y) - 2r(x, y) \sigma(x) \sigma(y)$. Thus, by Eq. 12, the form of the psychometric function is

$$p(x, y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{[u(x)-u(y)]/\sigma(x,y)} e^{-t^2/2}\, dt.$$

$$= \int_0^\infty N[u(x) - u(y), \sigma(x, y)],\tag{14}$$

where $N(\mu, \sigma)$ is the normal distribution with mean $\mu$ and standard deviation $\sigma$.

---

[8] The question of a correlation between stimulus effects arose a number of times in the discussion of detection (see especially Secs. 1, 2 and 6.2 of Chapter 3), and in all cases it was handled in one of two extreme ways. When the presentation consisted of the null stimulus followed by a tone (both in a background of noise), the correlation was assumed to be zero; when it was the null stimulus followed by an increment in a background such as a tone, the correlation was assumed to be perfect. One might anticipate that we would again postulate a correlation coefficient of one because the two stimuli of a discrimination experiment differ on only one physical dimension; however, that assumption, coupled with the decision rule given in Eq. 12, leads to perfect discrimination, which is contrary to fact. The reason that the assumption of perfect correlation worked in the analysis of detection but does not work here lies in the difference between the two psychophysical models that are assumed. There we used a threshold (quantal) model, whereas here we are assuming a continuous one. Presumably it is possible to develop a threshold model for discrimination, and it would be interesting to see how it differs from existing models.

Note that this model, as it stands, cannot handle the constant-error problem, for, when $x = y$, $[u(x) - u(y)]/\sigma(x, y) = 0$, and $p(x, x) = \frac{1}{2}$. This is because we have supposed that the order of presentation does not matter. A generalization that includes biases is given in Sec. 4.1.

If we let $Z(x, y)$ denote the normal deviate corresponding to $p(x, y)$, that is, $p(x, y) = \int_{-\infty}^{Z(x,y)} N(0, 1)$, then Eq. 14 can be written

$$u(x) - u(y) = Z(x, y)\,\sigma(x, y)$$
$$= Z(x, y)\{\sigma(x)^2 + \sigma(y)^2 - 2\sigma(x)\,\sigma(y)\,r(x, y)\}^{\frac{1}{2}}, \quad (15)$$

which is what Thurstone sometimes called the *law* and sometimes the *equation of comparative judgment*. Because its status as a law is far from certain, we prefer the more neutral word *equation*.

We notice, first, that the function $u$, which gives the means of the random variables associated with the stimuli, is a scale of the sort that interested Fechner. Second, the equally-often-noticed-difference problem has a solution if and only if the variances $\sigma(x, y)$ of the discriminal differences are all equal. Third, without some further assumptions, even with complete pair-comparison data, there are more unknowns than there are equations. With $k$ stimuli, there are $k - 1$ unknown means (fixing one arbitrarily determines the zero of the scale), $k - 1$ unknown variances (fixing one sets the unit), and, including the $\langle x, x \rangle$ presentations, $k(k + 1)/2$ unknown correlations, but there are only $k(k - 1)/2$ equations.

To cope with this indeterminancy, Thurstone singled out five special cases of increasingly strong assumptions, of which Case V is the most familiar. It assumes that the discriminal dispersions are all the same, $\sigma^2(x) = \sigma^2$ for all $x$, and that the correlations are all the same (usually assumed to be 0), $r(x, y) = r$. If we choose $\sigma = 1/\sqrt{2(1 - r)}$, as we may because it merely determines the unit of our scale, then for Case V the equation of comparative judgment, Eq. 15, reduces to just

$$u(x) - u(y) = Z(x, y). \quad (16)$$

Thus the data in the form of the deviates $Z(x, y)$ give the scale when one of the scale values is chosen arbitrarily. Of course, the system of equations is now overdetermined: it consists of $k(k - 1)/2$ linear equations in $k - 1$ unknowns. It is easy to see that this imposes severe internal constraints on the data, among them

$$\begin{aligned} Z(x, z) &= u(x) - u(z) \\ &= u(x) - u(y) + u(y) - u(z) \\ &= Z(x, y) + Z(y, z), \end{aligned} \quad (17)$$

which amounts to saying that, if $p(x, y)$ and $p(y, z)$ are given, then $p(x, z)$ is determined.

Methods for using all the data to determine scale values are described in Green (1954), Guilford (1954), and Torgerson (1958), and procedures for evaluating the adequacy of the model in Mosteller (1951a,b,c).

Suppose that the data satisfy the conditions of Case V. If Weber's law is true, then by Theorem 4 $u$ must be a logarithmic function of stimulus intensity, which suggests plotting the psychometric function as a function of the logarithm of intensity. This is frequently done, the independent variable being $10 \log_{10}$ of the relative energy—the so-called decibel (db) scale. The data are usually well approximated by a cumulative normal distribution, as assumed indirectly by Thurstone and as Fechner suggested in his work. The assumption that the psychometric function is a cumulative normal of the physical stimulus was referred to as the "phi-gamma hypothesis" by Urban (1907, 1910), but, as Boring (1924) pointed out, this is not really a hypothesis until the physical measure is specified.

Thurstone's psychophysical theory is like Fechner's in that an underlying scale plays a crucial role; however, in the general case, the scale does not meet the equal-jnd property central to Fechner's work. It is a more specific theory than Fechner's in that the form of the psychometric function is taken to be a cumulative normal distribution. Finally, in the important Case V, in which the two theories overlap, the observed data must satisfy the strong constraint given in Eq. 17.

### 3.2 The Choice Axiom

A second approach to these questions of the form of the psychometric function and a scale of "sensation" has recently been suggested (Luce, 1959). The tack is a bit different from Thurstone's in that one begins with an assumption about the choice probabilities and from that derives a scaling model. This model is extremely similar to that described in Sec. 1.2 of Chapter 3.

So far, we have confined our attention to presentations of stimulus pairs; however, it is clear that we can generalize our experiments in several ways so that three or more stimuli are presented. Suppose that $T \subseteq \mathscr{S}$ has $k$ stimuli and that $\sigma(T)$ is a simple ordering of $T$. Thus $\sigma(T) \in \mathscr{S} \times \mathscr{S} \times \ldots \times \mathscr{S}$ ($k$ times) and $\sigma(T)$ is a possible presentation of $k$ stimuli. By $\sigma_r(T)$ we denote the element of $T$ that is in the $r$th position of the $k$-tuple $\sigma(T)$. If $R = \{1, 2, \ldots, k\}$, then the identification function, which can be stated formally if one wishes, simply says that response $r$ means that the largest element is believed to have been in the $r$th position of the ordering.

In this section we are supposing that order of presentation does not matter, that is, if $\sigma_r(T) = \sigma_r'(T) = x$, then

$$p[r \mid \sigma(T)] = p[r' \mid \sigma'(T)].\qquad(18)$$

Because the response in both cases designates the same stimulus as the largest, it is again convenient to speak as if the subject made the choice $x$ from the unordered set $T$ and to denote the common value in Eq. 18 by $p_T(x)$. Note that the basic choice set $T$ is written as a subscript; this is done in order to simplify writing other conditional probabilities later. To conform to the previous notation, we denote $p_{\{x,y\}}(x)$ by $p(x, y)$.

The probability that the subject's choice is an element from the subset $U \subseteq T$ is given by

$$p_T(U) = \sum_{x \in U} p_T(x).$$

The usual probability axioms are assumed to hold for each $T$.

The question of empirical interest is what, if any, added relations that are not logical consequences of the probability axioms are imposed by the organism. One suspects that the probability measure associated with the presentation of one set $T$ cannot be totally independent of that governing the behavior when sets overlapping $T$ are presented—that the organism is unable or unwilling to generate totally unrelated choices in two related situations. If there are no such relations, a whole range of possible scientific questions has no significance; if there are relations, their statement is a piece of a theory of behavior.

It may well be that, although they exist, these relations are very complicated; however, it seems best to begin with the simplest conditions that are not completely at variance with known facts. The one suggested by Luce and, in another context, by Clarke (1957), states, in essence, that if some alternatives are removed from consideration then the relative frequency of choices among the remaining alternatives is preserved. Put another way, the presence or absence of an alternative is irrelevant to the relative probabilities of choice between two other alternatives, although, of course, the absolute value of these probabilities will generally be affected. To cast this in mathematical language, we need the definition of *conditional probability*, namely

$$\text{if } p_T(U) \neq 0, \quad \text{then } p_T(V \mid U) = \frac{p_T(U \cap V)}{p_T(U)}.$$

We assume the following:

**The Choice Axiom.**   *If $x \in U \subseteq T$ and if $p_T(x \mid U)$ exists, then*

$$p_T(x \mid U) = p_U(x).$$

We first prove Theorem 5 and then discuss its significance.

**Theorem 5.** *Suppose that the choice axiom holds for all* $U$, $x \in U \subseteq T$.

1. *If* $p_T(x) \neq 0$, *then* $p_U(x) \neq 0$.

2. *If* $p_T(x) = 0$ *and* $p_T(U) \neq 0$, *then* $p_U(x) = 0$.

3. *If* $p_T(y) = 0$ *and* $y \neq x$, *then* $p_T(x) = p_{T-\{y\}}(x)$.

4. *If, for all* $y \in T$, $p_T(y) \neq 0$, *then* $p_T(x) = p_U(x) \, p_T(U)$.

PROOF. 1. Because $\quad p_T(x) \neq 0, \quad p_T(U) = p_T(x) + \sum\limits_{y \in U - \{x\}} p_T(y) \neq 0$.

Therefore, the conditional probability $p_T(x \mid U) = p_T(x)/p_T(U)$ is defined, and it is not 0 because $p_T(x) \neq 0$. By the choice axiom, it equals $p_U(x)$.

2. If $p_T(x) = 0$ and $p_T(U) \neq 0$, then by the definition of conditional probability and by the choice axiom

$$p_U(x) = p_T(x \mid U) = \frac{p_T(x)}{p_T(U)} = 0.$$

3. Because $p_T(y) = 0$,

$$\begin{aligned} p_T(T - \{y\}) &= p_T(y) + \sum_{x \in T - \{y\}} p_T(x) \\ &= p_T(T) \\ &= 1. \end{aligned}$$

Using this,

$$\begin{aligned} p_T(x \mid T - \{y\}) &= \frac{p_T(\{x\} \cap [T - \{y\}])}{p_T(T - \{y\})} \\ &= p_T(x). \end{aligned}$$

So, by the choice axiom,

$$\begin{aligned} p_{T-\{y\}}(x) &= p_T(x \mid T - \{y\}) \\ &= p_T(x). \end{aligned}$$

4. By assumption, $p_T(y) \neq 0$ for $y \in U$, so $p_T(U) \neq 0$, and the choice axiom and the definition of conditional probability yield

$$p_U(x) = p_T(x \mid U)$$

$$= \frac{p_T(x)}{p_T(U)}.$$

Multiplying by $p_T(U)$ yields the result.

What does this tell us? If we have an alternative $y$ that is never chosen, Part 3 tells us that we may delete it from the set of alternatives. That we may keep repeating this process with no concern about the order in which it is done until all the choice probabilities are positive is guaranteed by Parts 1 and 2. The final part tells us how the various probabilities are

related when we have only positive probabilities. The next theorem makes this relation more vivid.

**Theorem 6.** *If, for all* $x \in T$, $p_T(x) \neq 0$ *and if the choice axiom holds for all* $x$ *and* $U$ *such that* $x \in U \subseteq T$, *then*

1. *for* $x$, $y$, $z \in T$,

$$p(x, z) = \frac{p(x, y)\, p(y, z)}{p(x, y)\, p(y, z) + p(z, y)\, p(y, x)} \tag{19}$$

and

2.       $$p_T(x) = \frac{1}{1 + \sum\limits_{y \in T-\{x\}} p(y, x)/p(x, y)}. \tag{20}$$

PROOF.   We begin by showing that

$$\frac{p(x, y)}{p(y, x)} = \frac{p_T(x)}{p_T(y)}. \tag{21}$$

By Part 4 of Theorem 5,

$$p_T(x) = p(x, y)p_T(\{x, y\})$$
$$= p(x, y)[p_T(x) + p_T(y)].$$

Rewrite this as

$$p_T(x)[1 - p(x, y)] = p(x, y)\, p_T(y),$$

and note that $p(y, x) = 1 - p(x, y)$. By Part 1 of Theorem 5 none of these probabilities is 0, so we may cross divide to get the assertion.

To prove Part 1, we note that

$$1 = \frac{p_T(x)\, p_T(y)\, p_T(z)}{p_T(y)\, p_T(z)\, p_T(x)}$$
$$= \frac{p(x, y)\, p(y, z)\, p(z, x)}{p(y, x)\, p(z, y)\, p(x, z)}. \tag{22}$$

Substitute $p(z, x) = 1 - p(x, z)$ and solve for $p(x, z)$ to get Eq. 19.

To prove Part 2, consider

$$1 + \sum_{y \in T-\{x\}} \frac{p(y, x)}{p(x, y)} = \frac{p_T(x)}{p_T(x)} + \sum_{y \in T-\{x\}} \frac{p_T(y)}{p_T(x)}$$
$$= \frac{1}{p_T(x)} \sum_{y \in T} p_T(y)$$
$$= \frac{1}{p_T(x)},$$

because $P_T(T) = 1$. Solve for $p_T(x)$ to get Eq. 20.

This excursion completed, let us now see what implications the choice axiom has for two-alternative discrimination. First, as with Thurstone's Case V, we find that there is a strong constraint on the psychometric

functions: if $p(x, y)$ and $p(y, z)$ are known, then $p(x, z)$ is determined by Eq. 19. It is obvious that Eq. 19 establishes a relation different from Thurstone's Eq. 17, which hints at an experiment to decide between the models. Unfortunately, a few calculations show that such an experiment is probably not practical. Table 2 presents representative predictions of the

Table 2    Comparison of Predicted $p(x, z)$ from Known $p(x, y)$ and $p(y, z)$ Using Choice Axiom and Thurstone's Case V

|  |  | $p(y, z)$ | | | |  | $p(y, z)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  | 0.6 | 0.7 | 0.8 | 0.9 |  | 0.6 | 0.7 | 0.8 | 0.9 |
|  | 0.6 | 0.692 | 0.778 | 0.857 | 0.931 | 0.6 | 0.695 | 0.782 | 0.864 | 0.938 |
| $p(x, y)$ | 0.7 |  | 0.845 | 0.903 | 0.954 | 0.7 |  | 0.853 | 0.915 | 0.965 |
|  | 0.8 |  |  | 0.941 | 0.973 | 0.8 |  |  | 0.954 | 0.983 |
|  | 0.9 |  |  |  | 0.988 | 0.9 |  |  |  | 0.995 |

$p(x, z)$ from the choice axiom       $p(x, z)$ from Thurstone's Case V

two models; the differences are simply not large enough to be detectable in practice. Presently, we shall see in another way how similar the two models are.

Part 2 of Theorem 6 suggests a way of introducing a scale into this model. To an arbitrary stimulus $s$ assign the scale value $v(s) = k$, where $k$ is some positive number. To any other stimulus $x$ assign the value $v(x) = kp(x, s)/p(s, x)$. Optimal procedures for estimating these $v$-scale values from data are discussed in Abelson & Bradley (1954), Bradley (1954a,b, 1955), Bradley & Terry (1952), Ford (1957). Now, consider

$$\frac{v(y)}{v(x)} = \frac{kp(y, s)\, p(s, x)}{p(s, y)\, kp(x, s)}.$$

Using Eq. 19 in the form of Eq. 22, we see that the right side can be replaced by $p(y, x)/p(x, y)$, so

$$\frac{v(y)}{v(x)} = \frac{p(y, x)}{p(x, y)}.\tag{23}$$

Substituting this into Eq. 21, we get the basic equation relating choice probabilities to the $v$-scale:

$$p_T(x) = \frac{1}{1 + \sum_{y \in T - \{x\}} v(y)/v(x)}$$

$$= \frac{v(x)}{\sum_{y \in T} v(y)}.\tag{24}$$

In the two-alternative case, this reduces simply to

$$p(x, y) = \frac{1}{1 + v(y)/v(x)}.$$

A number of authors (Bradley & Terry, 1952; Ford, 1957; Gulliksen, 1953; Thurstone, 1930) have proposed and studied this model. Now, if we define

$$u(x) = a \log v(x) + a' \tag{25}$$

then

$$p(x, y) = \frac{1}{1 + \exp\{[u(y) - u(x)]/a\}}, \tag{26}$$

which shows that the choice axiom is a sufficient condition to solve the revised Fechner problem. It also shows the form of the psychometric function in terms of the Fechnerian scale $u$; this is known as the logistic curve. It is well known that the logistic is quite a good approximation to the cumulative normal so again we see the close similarity of this model to Thurstone's Case V.

If the data satisfy the generalized Weber law (Eq. 6) and if $C(\pi)/K(\pi) = \gamma$, then we know by Theorem 4 that

$$u(x) = A \log (x + \gamma) + B. \tag{27}$$

Thus, solving for $v(x)$ by eliminating $u(x)$ from Eqs. 25 and 27, we obtain

$$v(x) = \alpha(x + \gamma)^{\beta},$$

where $\alpha$ and $\beta$ are constants. The former is a free, positive constant, but $\beta$ can be determined as follows. We know that

$$p(x, y) = \pi = \frac{1}{1 + [(y + \gamma)/(x + \gamma)]^{\beta}}$$

if and only if $x = y + K(\pi)(y + \gamma)$. Thus,

$$\beta = \frac{\log [(1 - \pi)/\pi]}{\log [(y + \gamma)/(x + \gamma)]}$$

and

$$\frac{y + \gamma}{x + \gamma} = \frac{1}{1 + K(\pi)}.$$

So,

$$\beta = \frac{\log [\pi/(1 - \pi)]}{\log [1 + K(\pi)]}.$$

Note that $\beta$ is a constant independent of $\pi$ and that this equation states how $K(\pi)$ varies with $\pi$ if the model is correct. Typically, for $\pi = 0.75$,

$K(\pi)$ is in the range for 0.01 to 0.10 (see Table 1, p. 205) in which case $\beta$ is in the range from 10 to 100.

It should be stressed again that this model, like Thurstone's, has CE = 0, for, when $x = y$, $p(x, y) = \frac{1}{2}$. In Sec. 4.2 a generalized choice model is presented in which nonzero CE's are possible.


### 3.3 Semiorders

The notion of a jnd, introduced in Sec. 1.2, is a summary of some of the information included in the psychometric function. It is an algebraic rather than probabilistic notion. Although nothing was said then about properties it might exhibit as an algebraic entity, they are of interest whether jnds are treated as behavioral data in their own right or as simple approximations to the psychometric functions.

One way to state such properties is to introduce two binary relations, their "boundary" coinciding with the notion of a $\pi$-jnd. For $\pi$, $\frac{1}{2} < \pi < 1$, define

$$
\begin{aligned}
xL(\pi)y \quad &\text{if and only if} \quad p(x, y) > \pi \\
xI(\pi)y \quad &\text{if and only if} \quad 1 - \pi \leqslant p(x, y) \leqslant \pi,
\end{aligned}
\tag{28}
$$

where $L$ is used to suggest "larger" and $I$, "indifference." If $xL(\pi)y$, then $x$ is more than one $\pi$-jnd larger than $y$, and, if $xI(\pi)y$, then $x$ is less than or equal to one $\pi$-jnd from $y$.

On a priori grounds, Luce (1956) suggested that such relations might satisfy the following axiom system:

*A pair of binary relations $I(\pi)$ and $L(\pi)$ over $\mathscr{S}$ form a semiordering of $\mathscr{S}$ if for all $x$, $y$, $z$, and $w \in \mathscr{S}$:*

1. *exactly one of $xI(\pi)y$, $xL(\pi)y$, or $yL(\pi)x$ holds;*
2. *$xI(\pi)x$;*
3. *when $xL(\pi)y$, $yI(\pi)z$, and $zL(\pi)w$, then $xL(\pi)w$;*
4. *when $xL(\pi)y$ and $yL(\pi)z$, then not both $xI(\pi)w$ and $wI(\pi)z$*

Some of the main features of a semiorder are much like those of a weak order: $L(\pi)$ is transitive and asymmetric and $I(\pi)$ is reflexive and symmetric. The important difference is that $I(\pi)$ need not be transitive. It is constrained, however, by Axioms 3 and 4 to have the feature that an indifference interval can never "span" a larger-than interval. For other discussions of semiorders, including an axiomatization in terms of the single relation $L$, see Scott & Suppes (1958), Gerlach (1957), and Sec. 3.2 of Chapter 1.

In Luce (1959) it is shown that if the probabilities satisfy the choice

model, then $L(\pi)$ and $I(\pi)$ as defined in Eq. 28 form a semiordering of $\mathscr{S}$. We shall not reproduce the proof here. Rather, we shall show that the same is true for Thurstone's Case V model. Let $Z(\pi)$ be defined by $\pi = \int_{-\infty}^{Z(\pi)} N(0, 1)$, where $N(0, 1)$ is the normal distribution with zero mean and unit variance. By Eqs. 14 and 28,

$$xL(\pi)y \quad \text{if and only if} \quad u(x) - u(y) > \sigma Z(\pi)$$

$$xI(\pi)y \quad \text{if and only if} \quad -\sigma Z(\pi) \leqslant u(x) - u(y) \leqslant \sigma Z(\pi).$$

It is obvious that properties 1 and 2 of a semiorder are met. If $xL(\pi)y$, $yI(\pi)z$, and $zL(\pi)w$, then $u(x) - u(y) > \sigma Z(\pi)$, $u(y) - u(z) \geqslant -\sigma Z(\pi)$, and $u(z) - u(w) > \sigma Z(\pi)$. So $u(x) - u(w) = u(x) - u(y) + u(y) - u(z) + u(z) - u(w) > \sigma Z(\pi) - \sigma Z(\pi) + \sigma Z(\pi) = \sigma Z(\pi)$, which implies $xL(\pi)w$. Thus condition 3 is met. Suppose condition 4 is false, that is, for some $x$, $y$, $z$, $w \in \mathscr{S}$, $xL(\pi)y$, $yL(\pi)z$, $xI(\pi)w$, and $wI(\pi)z$, then $u(x) - u(y) > \sigma Z(\pi)$ and $u(y) - u(z) > \sigma Z(\pi)$ imply $u(x) - u(z) > 2\sigma Z(\pi)$. But $u(x) - u(w) \leqslant \sigma Z(\pi)$ and $u(w) - u(z) \leqslant \sigma Z(\pi)$ imply $u(x) - u(z) \leqslant 2\sigma Z(\pi)$. This contradiction shows that condition 4 is met.

For the more general Thurstone models, the relations defined by Eqs. 28 need not satisfy the semiorder axioms.

## 4. BIASED RESPONSE MODELS

It is a curious historical fact that the CE has not received nearly the attention that the jnd has. Possibly, this is because the jnd is, in fact, a measure of discriminability, whereas, from the point of view of one interested in discrimination, the CE is after all a nuisance factor—albeit an ubiquitous one. Moreover, because psychologists have by their designs and procedures tried to achieve maximum discriminability and have not been concerned how the jnd varies with experimental manipulations, there has really been only the single question: how does the jnd depend upon the physical measures of the stimuli under study? With the CE, matters are not so simple, and in the past interest seems not to have been great. In the last few years, however, theorists have begun to give it more attention, and we anticipate increased experimental interest. There is, for example, the result of Eq. 7 in Sec. 1.4, which suggests that the CE may follow Weber's generalized law when the jnd does; this needs to be tested, but more interesting are questions that can be raised about the impact of payoffs and other experimental variables on the CE (see Sec. 4.3). As early as 1920, Thompson discussed this problem, but evidently this did not lead to empirical research. Boring (1920) expressed the

attitude of the time that these are matters for experimental control, not investigation.

Whether or not we study the CE as such, its mere existence shows that the two response models just discussed cannot be correct and suggests that we introduce response biases in some fashion. We now look into this question.

## 4.1 The Equation of Comparative Judgment

The most obvious way to introduce response biases into Thurstone's model is to modify the decision rule, Eq. 12. We do so as follows:

*When $\langle x, y \rangle$ is presented, the first stimulus, $x$, is judged $\begin{Bmatrix} larger \\ smaller \end{Bmatrix}$ than the second, $y$, if $\mathbf{X} \begin{Bmatrix} > \\ < \end{Bmatrix} \mathbf{Y} + c$, where $c$ is some positive or negative number.* $\hspace{2cm}$ (29)

With this rule, we have, as before,

$$
\begin{aligned}
p(1 \mid \langle x, y \rangle) &= \Pr\left(\mathbf{X} > \mathbf{Y} + c\right) \\
&= \Pr\left(\mathbf{X} - \mathbf{Y} - c > 0\right) \\
&= \int_{-\infty}^{\frac{u(x)-u(y)-c}{\sigma(x,y)}} N(0,\ 1) \\
&= \int_{c}^{\infty} N[u(x) - u(y), \sigma(x, y)],
\end{aligned}
\tag{30}
$$

where $N(\mu, \sigma)$ is the normal distribution with mean $\mu$ and standard deviation $\sigma$. This modification of Thurstone's model was first suggested by Harris (1957). (If we had simply assumed an additive bias so that the mean of the effect of stimulus $x$ is $u(x) + a(1)$ when $x$ is presented first and $u(x) + a(2)$ when it is presented second, then setting $c = a(2) - a(1)$ yields Eq. 30. As we shall see in Eq. 32, this means that the way biases are introduced into the choice model is substantially the same as this way of introducing them into Thurstone's model, except for a logarithmic transformation.)

Consider a pair-comparison design involving $k$ stimuli. The unknowns are $c$, the $k$ means $u(x)$, the $k$ standard deviations, $\sigma(k)$, and the $k(k + 1)/2$ correlations, $r(x, y)$, a total of $1 + k + k + k(k + 1)/2 = 1 + k(k + 5)/2$ variables. If we set one of the means equal to 0 and one of the standard deviations equal to 1, which only fixes the zero and unit of the scale, this reduces the number of unknowns to $k(k + 5)/2 - 1$. There are, of course,

$k^2$ equations. The number of unknowns is easily seen to be less than or equal to the number of equations for $k \geqslant 5$, and if we are willing to introduce the constraints of the form

$$[u(x) - u(y)] + [u(y) - u(z)] = u(x) - u(z)$$

as assumptions rather than use them to test the model then the minimum number of stimuli needed to estimate the parameters is reduced to 4.

Thus, introducing the cutoff $c$ not only makes the model more realistic but, in principle, at least, renders the model determinant, which we recall it was not in Thurstone's original formulation. Just how to solve for the unknowns appears, however, to be an extremely formidable problem, involving as it does 25 unknowns in 25 nonlinear equations.

Certain special cases are amenable to solution and, as in Thurstone's work, may serve as useful first approximations. For example, suppose $r(x, y) = 0$ for all $x$ and $y$. For some particular stimulus $s$, we may arbitrarily set $u(s) = 0$ and $\sigma(s) = 1/\sqrt{2}$, and equating the relative frequencies of choices to the theoretical probabilities we estimate $c$ from

$$p(1 \mid \langle s, s \rangle) = \int_{\hat{c}}^{\infty} N(0, 1).$$

With $\hat{c}$ so determined, we estimate $\sigma(x)$ from

$$p(1 \mid \langle x, x \rangle) = \int_{\hat{c}}^{\infty} N[0, \sqrt{2}\, \hat{\sigma}(x)].$$

Finally, with $\hat{c}$ and $\hat{\sigma}(x)$ known for all $x$, we estimate $u(x) - u(y)$ from

$$p(1 \mid \langle x, y \rangle) = \int_{\hat{c}}^{\infty} N\{\hat{u}(x) - \hat{u}(y), [\hat{\sigma}(x)^2 + \hat{\sigma}(y)^2]^{1/2}\}.$$

There are various internal checks on the adequacy of the model, including the two estimates of $u(x) - u(y)$ from $p(1 \mid \langle x, y \rangle)$ and from $p(1 \mid \langle y, x \rangle)$ and the additivity condition on triples of stimuli mentioned earlier.

As far as we know, no empirical research has been carried out using this generalization of the Thurstone model in the context of discrimination, although for a similar model in detection there are considerable data (see Chapter 3).

It follows immediately from Eq. 30 that $p(1 \mid \langle s, x_{1/2} \rangle) = \frac{1}{2}$ if and only if $u(s) - u(x_{1/2}) - c = 0$. If we suppose that Fechner's law is correct so that $u(s) = A \log (s + \gamma) + B$ and define $b$ so that $c = -A \log b$, then

$$A \log (s + \gamma) + B - A \log (x_{1/2} + \gamma) - B + A \log b = 0$$

implies $x_{1/2} = b(s + \gamma) - \gamma$. Thus, $CE = x_{1/2} - s = (b - 1)(s + \gamma)$, which again is the generalized Weber law for the CE.

## 4.2 A Choice Model

A generalization of the unbiased choice model (Sec. 3.2), but not of the choice axiom itself, was described by Luce (1959). Although the argument leading to the model is somewhat indirect and to some people not entirely satisfactory, the model itself is easily stated. Let $s = \langle \varDelta_1, \varDelta_2, \ldots, \varDelta_k \rangle$ $\in S$, where $\varDelta_r \in \mathscr{S}$, and $R = \{1, 2, \ldots, k\}$, then there are two numerical ratio scales

$$\eta : \mathscr{S} \to \text{positive reals}$$
$$b : R \to \text{positive reals}$$

such that

$$p(r \mid s) = \frac{\eta(\varDelta_r)\, b(r)}{\displaystyle\sum_{r' \in R} \eta(\varDelta_{r'})\, b(r')}. \tag{31}$$

This representation is closely similar to that given in Eq. 24, except that now the order of presentation of the stimuli matters. Both the stimulus and the response used to designate it contribute multiplicatively and independently to the response probability in Eq. 31. It is appropriate to think of $\eta(\varDelta)$ as a stimulus parameter associated with stimulus $\varDelta$, which is independent of the response used to designate it, and to think of $b(r)$ as a response parameter, or bias, associated with the response, which is independent of the stimulus to which the response refers. If there is no differential tendency to use the responses, that is, if $b(r)$ is a constant independent of $r \in R$, then the $b$ terms drop out and Eq. 31 reduces to Eq. 24.

This representation is also similar to the choice model for complete identification experiments described in Sec. 1.2 of Chapter 3. The only formal difference is that the stimulus parameters here depend upon single stimuli, whereas there they depend upon pairs of stimuli. This difference may well be more apparent than real, as we shall see shortly.

One way to arrive at Eq. 31 is via a learning argument similar to that given in Sec. 1.2 of Chapter 3 (see Bush, Luce, & Rose, 1963). Consider the special but important design in which $\mathscr{S}$ has $k$ elements and each presentation includes every element of $\mathscr{S}$. Put another way, the presentations are just orderings of $\mathscr{S}$. Let $\varDelta^* \in \mathscr{S}$ denote the "correct" element— the one meeting the discriminative criterion. In terms of the identification function, $\iota(r) = \{s \mid \varDelta_r = \varDelta^*\}$. Suppose that $s'$ is presented and $s' \in \iota(r')$; then the response probabilities are assumed to be transformed linearly by

$$p_{n+1}(r \mid s) = p_n(r \mid s) + \eta(\varDelta_{r'}, \varDelta^*)\, \theta(r')[\delta_{rr'} - p_n(r \mid s)].$$

As in the complete identification situation, $\theta$ is a basic learning rate parameter associated with the correct response and $\eta$ is a generalization parameter. Note that by our assumptions the generalization is not between presentations but between the stimuli corresponding to the correct response, one of these stimuli, of course, being $\varDelta^*$.

Paralleling the argument used in Sec. 1.2 of Chapter 3, it is easy to show that

$$\lim_{n \to \infty} E[p_n(r \mid s)] = \frac{\eta(\varDelta_r, \varDelta^*)\, b(r)}{\sum_{r' \in R} \eta(\varDelta_{r'}, \varDelta^*)\, b(r')} \,,$$

where $b(r) = P(r)\,\theta(r)$, and $P(r)$ denotes the a priori probability that $r$ is the correct response, that is, $P(r)$ is the sum of all the presentation probabilities $P(s)$ for which $s \in \iota(r)$. This equation has the same form as Eq. 31, except that $\eta(\varDelta_r)$ is replaced by $\eta(\varDelta_r, \varDelta^*)$. Because $\varDelta^*$ is fixed throughout, this is a purely notational difference. Thus, the discrimination choice model for at least this special case is not as different from the complete identification model as it first seems.

If one does not assume that the presentations are orderings of $\mathscr{S}$, then different stimuli are correct for different presentations and matters are more complicated. In particular, the asymptotic form is not Eq. 31, and so, for example, the analysis of the method of constant stimuli given below in terms of Eq. 31 may very well be wrong.

It is not difficult to see that Eq. 31 implies: for any $s, s' \in S$ and any $r, r', r'' \in R$ such that $\varDelta_r = \varDelta_{r'}'$, then $[p(r \mid s)/p(r'' \mid s)][p(r'' \mid s')/p(r' \mid s')]$ is a function of only $r$, $r'$, $\varDelta_{r''}$, and $\varDelta_{r''}'$. It has been shown that this property, which involves only observable quantities, is in fact equivalent to Eq. 31 (Luce, 1962).

The empirical adequacy of Eq. 31 is readily tested because the number of parameters increases linearly with the number of stimuli used and with the number of response categories, whereas the number of probabilities to be accounted for increases much more rapidly. For example, if three stimuli are used, they may be presented in six different orders when the subject is to select one out of three and also in six ordered pairs when he is to select one out of two, yielding a total of 18 independent conditional probabilities. Because we can, and must, arbitrarily choose the unit of the $\eta$ and $b$ scales in Eq. 31, there are two, not three, stimulus parameters and, assuming different biases in the two- and three-choice situations, a total of three bias parameters. So there are $18 - 5 = 13$ degrees of freedom even in this simple situation.

Such an experiment was performed on lifted weights by Shipley and Luce (1963) and on visual brightness by van Laer (undated). As an example of
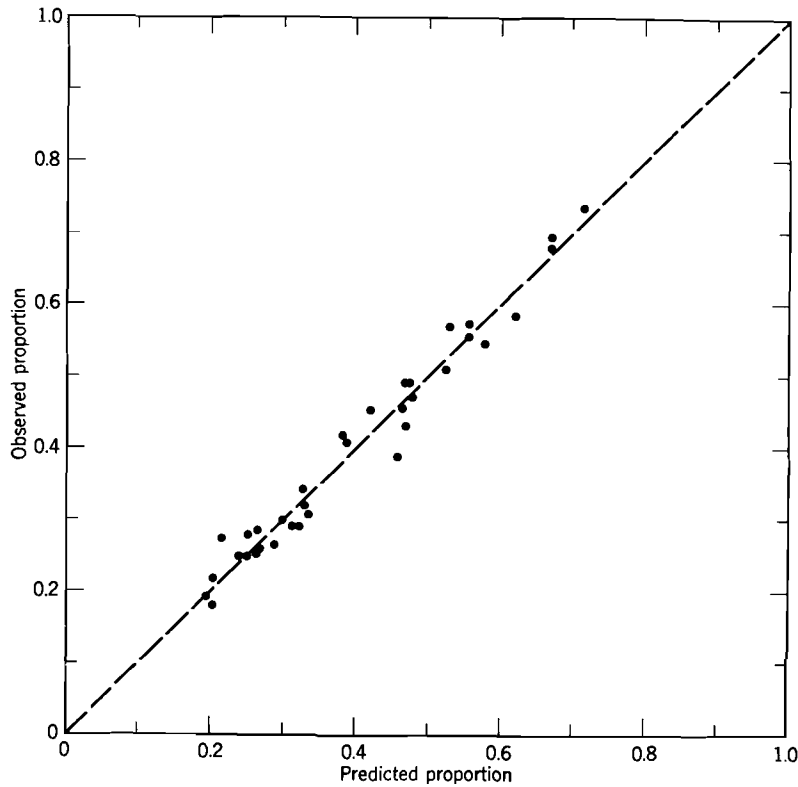
Fig. 10. Observed versus expected proportions predicted by the choice model for lifted weight data collected under two procedures. In one, each of the six orderings of three weights plus three repetitions of the intermediate weight were presented 400 times. In the other, each of the six orderings of two of the three weights plus two repetitions of the intermediate weight were presented 400 times. Stimulus and bias parameters were estimated from the data for each condition using a technique suggested in Luce (1959). The geometric means of the two estimates of the stimulus parameters were used to calculate the expected proportions. For this subject, $\chi^2 = 24.00$, df $= 16$, and $0.05 < p < 0.10$. These data were collected by Shipley and Luce (1963).

the adequacy of the model, Fig. 10 shows the predicted versus observed proportions for one subject in the weight-lifting experiment.

Consider, now, the method of constant stimuli, where the standard stimulus $s$ is presented first, the comparison $x$, second, and the response categories are 1 and 2. Then

$$p(1 \mid \langle s, x \rangle) = \frac{1}{1 + \eta(x)/\eta(s)b}, \tag{32}$$

where $b = b(1)/b(2)$. If we let $u(s) = a \log \eta(s) + a'$ and $c = -a \log b$, then it is easy to see that Eq. 32 can be rewritten in the logistic form

$$p(1 \mid \langle s, x \rangle) = \frac{1}{1 + \exp \{[u(x) - u(s) + c]/a\}}.$$

Questions of estimating the parameters $a$ and $c$ from data are discussed in Chapter 8.

By setting $p(1 \mid \langle s, x_{1/2} \rangle) = p(2 \mid \langle s, x_{1/2} \rangle) = \frac{1}{2}$, it is easy to see that $\eta(x_{1/2}) = b\eta(s)$. If the $\eta$-scale is a power function, $\eta(x) = \alpha(x + \gamma)^\beta$, then

$$x_{1/2} = b^{1/\beta}(s + \gamma) - \gamma,$$

so

$$\mathrm{CE} = x_{1/2} - s = (b^{1/\beta} - 1)(s + \gamma),$$

which is again the generalized Weber law for the CE. A similar calculation with $p[2 \mid \langle s, s+ \Delta(s, \pi) \rangle] = \pi$ yields

$$\Delta(s, \pi) = \left[ \left( \frac{b\pi}{1 - \pi} \right)^{1/\beta} - 1 \right] (s + \gamma),$$

and so

$$\pi\text{-jnd} = (\tfrac{1}{2})[\Delta(s, \pi) - \Delta(s, 1 - \pi)]$$

$$= (\tfrac{1}{2})(s + \gamma)b^{1/\beta}\left[ \left( \frac{\pi}{1 - \pi} \right)^{1/\beta} + \left( \frac{1 - \pi}{\pi} \right)^{1/\beta} \right],$$

which is a generalized Weber law.

Note that both the CE and the jnd depend upon the factor $b^{1/\beta}$ but that the jnd also depends upon a factor involving only $\beta$, namely $[\pi/(1 - \pi)]^{1/\beta} + [(1 - \pi)/\pi]^{1/\beta}$. It is really only the second term that measures discriminability if this model is correct. This suggests considering the quantity

$$\frac{\text{jnd}}{x_{1/2} + \gamma} = \frac{1}{2} \left[ \left( \frac{\pi}{1 - \pi} \right)^{1/\beta} + \left( \frac{1 - \pi}{\pi} \right)^{1/\beta} \right]$$

as a measure of discrimination rather than the Weber fraction jnd/$(s + \gamma)$, which depends upon $b$.

## 4.3 Payoffs

Little or no experimental work has been carried out to determine the effects of payoffs in discrimination designs. Some workers have used symmetric payoff matrices in an attempt to motivate the subjects to reduce the size of the jnd, but in contrast to the extensive use of payoffs to

manipulate the response probabilities in detection research, the effects of payoffs upon the psychometric function have not been investigated. It is difficult not to believe that analogous manipulations are possible and that, in particular, the magnitude and sign of the CE are a function of the payoffs used. All we can do at present is to carry out some theoretical calculations, which we do for the choice model. The parallel calculations for the Thurstone model are possible, but they are somewhat messier in detail.

Suppose we consider a constant stimulus experiment with the standard $s$ and variable stimuli $x_{-k} < x_{-k+1} < \ldots < x_{-1} < s < x_1 < \ldots < x_{k-1} < x_k$. Suppose that each presentation $\langle s, x_i \rangle$, $i = 0, \pm 1, \ldots, \pm k$, where $s = x_0$, occurs equally often, that is, with probability $1/(2k + 1)$. The subject is to say whether the first or second stimulus is larger, and after each trial he is paid off according to the following monetary payoff matrix:

$$
\begin{array}{cc}
& \begin{array}{cc} 1 & \quad 2 \end{array} \\
\begin{array}{l}
\text{Relation between} \quad x_i < s \\
\text{Stimuli in the} \quad\;\; x_i = s \\
\text{Presentation} \quad\;\;\; x_i > s
\end{array}
&
\left[
\begin{array}{cc}
o_{11} & o_{12} \\
0 & 0 \\
o_{21} & o_{22}
\end{array}
\right],
\end{array}
$$

where $o_{11} > o_{12}$ and $o_{22} > o_{21}$.
The expected monetary return on a typical trial is

$$
E(o) = \frac{1}{2k + 1} \left\{ \sum_{i=-1}^{-k} [o_{11}p(1 \mid \langle s, x_i \rangle) + o_{12}p(2 \mid \langle s, x_i \rangle)] \right.
$$

$$
\left. + 0 + \sum_{i=1}^{k} [o_{21}p(1 \mid \langle s, x_i \rangle) + o_{22}p(2 \mid \langle s, x_i \rangle)] \right\}.
$$

If the choice model is correct, then Eq. 32 applies. Writing $\eta_i$ for $\eta(s)/\eta(x_i)$, we see that

$$
E(o) = \frac{1}{2k + 1} \left\{ \sum_{i=-1}^{-k} \left[ \frac{o_{11}b\eta_i + o_{12}}{1 + b\eta_i} \right] + 0 \right.
$$

$$
\left. + \sum_{i=1}^{k} \left[ \frac{o_{21}b\eta_i + o_{22}}{1 + b\eta_i} \right] \right\}.
$$

Assuming that the subject selects the value of the bias parameter to maximize $E(o)$, we calculate the derivative of $E(o)$ with respect to $b$ and set it equal to 0:

$$
\frac{dE(o)}{db} = 0 = \frac{1}{2k + 1} \left\{ \sum_{i=-1}^{-k} \left[ \frac{(o_{11} - o_{12})\eta_i}{(1 + b\eta_i)^2} \right] + \sum_{i=1}^{k} \left[ \frac{(o_{21} - o_{22})\eta_i}{(1 + b\eta_i)^2} \right] \right\}. \quad (33)
$$

If the payoff matrix is symmetric in the sense that $o_{11} = o_{22}$ and $o_{12} = o_{21}$, then

$$\sum_{i=1}^{k} \frac{\eta_i}{(1 + b\eta_i)^2} = \sum_{i=-1}^{-k} \frac{\eta_i}{(1 + b\eta_i)^2} = \sum_{i=1}^{k} \frac{\eta_{-i}}{(1 + b\eta_{-i})^2} \, . \tag{34}$$

In general this equation is satisfied only if $b \neq 1$. Thus, even with symmetric payoffs, a bias may be needed to optimize the expected payoff. An important exception is when the stimuli are located symmetrically around $s$ in the sense that $\eta_{-i} = 1/\eta_i$ or in terms of log $\eta$ when log $s -$ log $\eta(x_{-i}) = \log \eta(x_i) - \log s$. If this is so and if $b = 1$, then

$$\frac{\eta_{-i}}{(1 + \eta_{-i})^2} = \frac{1/\eta_i}{(1 + 1/\eta_i)^2} = \frac{\eta_i}{(\eta_i + 1)^2} \, ,$$

and so Eq. 34 is satisfied.

When the pairs are not presented equally often, when the $o_{ij}$ must be treated as utilities that are nonlinear with money so that the payoff matrix is not really symmetric even if the money outcomes are, or when the stimuli are not symmetric about the standard in the sense that $\eta(x_{-i})/\eta(s) = \eta(s)/\eta(x_i)$, in all these cases the expected payoff is a maximum only if a bias is introduced. Thus it may not be surprising that we so often find nonzero CE's. The empirical question remains whether or not the CE varies as predicted when we use different payoff matrices. It should not be forgotten that some detection studies cast doubt upon the maximization-of-expected-value model, and, if it is inadequate there, probably it is inadequate here as well.

## 5. UNORDERED DISCRIMINATION

The responses used in what we may call an *unordered discrimination experiment* are like those ordinarily used when speaking of discrimination. The subject simply reports whether he believes the pair of stimuli presented to be different, without specifying just how they differ. Aside from its more specific responses, any forced-choice discrimination design is equally suited to the study of unordered discrimination, but the converse is not true. Same-different responses make sense for many pairs of stimuli for which larger-smaller judgments are meaningless: for example, color patches that differ in hue, brightness, and saturation or, more generally, any stimuli that differ on two or more physical variables. In this section we shall look into models that have been proposed for studies of this type.

## 5.1 The Matching Experiment

In what we call a matching experiment, the set $\mathscr{S}$ of stimuli can be anything that we can control with an acceptable error. Because they need not be physically ordered, as in the forced-choice discrimination design, a whole new range of possibilities opens up, including stimuli that we cannot easily characterize in detail as physical objects: photographs, samples of handwriting, patches of textured material, etc.

The set $S$ of stimulus presentations is usually a subset of $\mathscr{S} \times \mathscr{S}$, although the possibility of presenting $k$-tuples of stimuli certainly exists. Because all existing research concerns pairs, we shall confine our attention to them. As with forced-choice discrimination, we can use either the method of pair comparisons, where $S = \mathscr{S} \times \mathscr{S}$, or methods (e.g., constant stimuli) in which $S$ is a proper subset of $\mathscr{S} \times \mathscr{S}$.

The response set $R$ consists of two elements, "match" and "not match," "same" and "different," or some equivalent terms, which we denote by $M$ and $\bar{M}$, respectively.

To get at the identification function, we must first consider ways in which terms such as "same" and "different" might be used, for there is an inherent ambiguity in their meanings. Certainly when a subject says two stimuli are the same he cannot mean that they are strictly the same in all respects, for two things cannot occur in the same place at the same time and still be designated by the subject. Just how much "sameness" is required differs from study to study, ranging from the stringent demand that the two stimuli be identical in all measurable aspects, except for their space-time location, to the weaker requirement that they be the same with respect to some one physical measure, such as sound energy, to the considerably weaker request that they be of the same "subjective loudness." The last of these examples differs from the other two in that the experimenter has no independent measure with which to decide whether the subject's response is correct or incorrect; that is, there is no identification function.

We shall confine our attention to stimuli such that there is an equivalence relation $\sim$ defined over them in terms of their physical properties. In some experiments it may denote identity; in others, equivalence with respect to sound energy or the like. This defines the partition

$$I_M\ (S) = \{\langle x, y \rangle \mid \langle x, y \rangle \in S \text{ and } x \sim y\}$$
$$I_{\bar{M}}\ (S) = S - I_M(S),$$

hence the identification function is

$$\iota(M) = I_M(S)$$
$$\iota(\bar{M}) = I_{\bar{M}}(S).$$

We shall suppose that the payoff matrix is compatible with the identification function in the sense that for $\langle x, y \rangle \in S$, it is

$$
\begin{array}{cc}
 & \text{Response} \\
 & M \quad \bar{M} \\
\begin{array}{c} x \sim y \\ \text{not } (x \sim y) \end{array} &
\begin{bmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{bmatrix},
\end{array}
$$

where $o_{11} > o_{12}$ and $o_{22} > o_{21}$.

The final ingredient in the design of a matching experiment is the schedule that assigns a presentation to each trial. As is customary in most of psychophysics, the assignment is random; however, equal probabilities for the several presentations are not generally used. The reason is simple. In a pair-comparison design with 10 stimuli and a criterion of strict identity there are 10 presentations, the $\langle x, x \rangle$ ones, for which "match" is the correct response, and 90 for which it is wrong. Thus, if the presentations were equiprobable, the subject would be wise to bias his responses strongly, possibly exclusively, to "not match." To avoid this bias, one can either increase the frequency of the $\langle x, x \rangle$ presentations, decrease the number of different $\langle x, y \rangle$ pairs presented, or use some mixture of these two that results in each response being correct half the time. If the outcomes are money and symmetric in the sense that $o_{11} = -o_{12} = -o_{21} = o_{22}$ and the subject responds entirely at random, then his expected outcome is zero for such schedules.

## 5.2 A Thurstonian Model

As far as we know, no probability model has been stated in the literature for the matching experiment. We do not see how to extend the choice model to unordered discrimination, but as it is easy to do so for the Thurstone model, we include that.

As in the forced-choice discrimination model, we suppose that to each stimulus $x \in \mathscr{S}$ there is a random variable $X$ assuming values on the real line; it is again interpreted as the momentary effect of the presentation of $x$. We assume that $X$ is normally distributed with mean $u(x)$ and standard deviation $\sigma(x)$ and that, when $x$ and $y$ are presented successively, the correlation between $X$ and $Y$ is $r(x, y)$. In forced-choice discrimination the choice between $x$ and $y$ was made by comparing $X - Y$ with some fixed cutoff $c$ (Sec. 4.1). Here the decision is between saying that the stimuli are the same or that they are different, and the plausible decision rule is that the subject selects two cutoffs, $-c$ and $d$, where $-c < d$, and that

$$
he\ responds\ \begin{Bmatrix} M \\ \bar{M} \end{Bmatrix} \quad if \quad \begin{Bmatrix} -c \leqslant X - Y \leqslant d \\ otherwise \end{Bmatrix}. \tag{35}
$$

In other words, if the effects are sufficiently close, he says the stimuli do not differ; otherwise, he says they do.

Because the distribution of the difference $X - Y$ is normal, with mean $u(x) - u(y)$ and standard deviation,

$$\sigma(x, y) = [\sigma^2(x) + \sigma^2(y) - 2\sigma(x)\,\sigma(y)\,r(x, y)]^{1/2},$$

the decision rule yields the representation

$$p(M \mid \langle x, y \rangle) = \int_{-c}^{d} N[u(x) - u(y), \sigma(x, y)],\qquad(36)$$

where $N(\mu, \sigma)$ denotes the normal distribution with mean $\mu$ and standard deviation $\sigma$.

Note that the often assumed, but probably incorrect, property of response symmetry $p(M \mid \langle x, y \rangle) = p(M \mid \langle y, x \rangle)$ is equivalent to $d = c$.

An equations-and-unknowns count similar to that carried out in Sec. 4.1 yields the same number of equations and one more unknown, still leaving five as the smallest number of stimuli for which the number of equations is not less than the number of unknowns. Again, nothing is known about solving this general case.

Previously we were able to solve the special case $r(x, y) = 0$, for all $x$ and $y$, by setting $u(s) = 0$ and $\sigma(s) = 1/\sqrt{2}$. Doing this here, we obtain from Eq. 36

$$p(M \mid \langle s, s \rangle) = \int_{-c}^{d} N(0, 1),$$

$$p(M \mid \langle x, x \rangle) = \int_{-c}^{d} N[0, \sqrt{2}\,\sigma(x)],$$

$$p(M \mid \langle x, s \rangle) = \int_{-c}^{d} N\{u(x), [\tfrac{1}{2} + \sigma^2(x)]^{1/2}\},$$

$$p(M \mid \langle s, x \rangle) = \int_{-c}^{d} N\{-\,u(x), [\tfrac{1}{2} + \sigma^2(x)]^{1/2}\},$$

which is a system of four equations in four unknowns, $c$, $d$, $u(x)$, and $\sigma(x)$. Although we do not know how to solve this system, mainly because there are two unknowns—$c$ and $d$—in the first equation, at least the number is sufficiently small so that one can hope that usable techniques for calculating the solutions can be found.

If we assume both $r(x, y) = 0$, for all $x$ and $y$, and that $d = c$, then solutions are easy to find:

$$p(M \mid \langle s, s \rangle) = \int_{-c}^{c} N(0, 1)$$

determines $c$,

$$p(M \mid \langle x, x \rangle) = \int_{-c}^{c} N[0, \sqrt{2}\, \sigma(x)]$$

determines $\sigma(x)$, a similar equation determines $\sigma(y)$, and

$$p(M \mid \langle x, y \rangle) = \int_{-c}^{c} N\{u(x) - u(y), [\sigma^2(x) + \sigma^2(y)]^{1/2}\}$$

determines $u(x) - u(y)$.

To get some idea about the effects of payoffs in this model, consider the simplest experiment and the simplest model. The experiment involves the presentation of either $\langle s, s \rangle$ with probability $P$ or $\langle s, s' \rangle$ with probability $1 - P$, and the payoff matrix is

$$\begin{array}{cc} & \text{Response} \\ & M \quad \bar{M} \end{array}$$

$$\begin{array}{c} \text{Stimulus} \\ \text{Presentation} \end{array} \quad \begin{array}{c} \langle s, s \rangle \\ \langle s, s' \rangle \end{array} \begin{bmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{bmatrix}.$$

Thus the expected outcome is

$$E(o) = [p(M \mid \langle s, s \rangle) - \beta p(M \mid \langle s, s' \rangle)]P(o_{11} - o_{12}) + Po_{12} + (1 - P)o_{22},$$

where

$$\beta = \left(\frac{1 - P}{P}\right)\left(\frac{o_{22} - o_{21}}{o_{11} - o_{12}}\right).$$

The model is the one just stated in which $d = c$, with the added assumption that all of the variances are equal, so that the variance of the difference distribution may be taken to be 1. Differentiating $E(o)$ with respect to $c$ and setting that equal to zero yields

$$\frac{dE(o)}{dc} = 0 = \frac{dp(M \mid \langle s, s \rangle)}{dc} - \beta\frac{dp(M \mid \langle s, s' \rangle)}{dc}$$

$$= 2\exp\left(\frac{-c^2}{2}\right) - \beta\left(\exp\left\{\frac{-[u(s) - u(s') - c]^2}{2}\right\}\right.$$

$$\left. + \exp\left\{\frac{-[u(s) - u(s') + c]^2}{2}\right\}\right)$$

$$= 2\exp\left(\frac{-c^2}{2}\right) - \beta 2\exp\left\{\frac{-[u(s) - u(s')]^2}{2}\right\}$$

$$\times \exp\left(\frac{-c^2}{2}\right)\cosh[u(s) - u(s')]c$$

$$= 2\exp\left(\frac{-c^2}{2}\right)\left(1 - \beta\exp\left\{\frac{-[u(s) - u(s')]^2}{2}\right\}\cosh[u(s) - u(s')]c\right).$$

So

$$\cosh [u(s) - u(s')]c = \frac{1}{\beta} \exp \left\{ \frac{[u(s) - u(s')]^2}{2} \right\} .$$

This last equation, which determines $c$, depends not only upon the payoff conditions through the parameter $\beta$ but also on the spacing of the stimuli $u(s) - u(s')$. In more complicated experiments the dependence upon the spacing of the stimuli remains but is, of course, more complex. It is evident from these considerations that the subject's definition of the match category is highly sensitive to the distribution of the stimuli, and one would guess that it is likely to be unstable in a given experiment unless the subject is well practiced. This may well be the theoretical counterpart of the empirical observation that indifference categories are very tricky to use and that, on the whole, it is better in discrimination work to use forced-choice procedures. (See Sec. 7.1 of Chapter 3 for a caution about extending this recommendation outside discrimination work.)

It appears that a series of experiments needs to be performed on the same subjects and with the same stimuli in which both forced-choice and unordered discrimination designs are used and the payoffs varied. The data should then be analyzed in terms of the Thurstonian models which, if they are correct, will yield approximately the same estimates of $u$ and $\sigma$ from all conditions. If that is verified, then the dependence of the cutoffs on the payoffs should be investigated, for maximization of expected money is not obviously the correct decision mechanism.

## 5.3 An Algebraic Model

In Chapter X of *The Structure of Appearance* Goodman (1951) presented an algebraic model for unordered discrimination. This model was later restated in slightly modified form and applied to data in visual psychophysics by Galanter (1956). Both presentations used the language of symbolic logic; we restate it in set terminology.

Goodman's work was not motivated by psychophysics, as such, but by philosophical considerations. To what extent can an ordering of stimuli be constructed from what in Goodman's view is the most primitive observing response: matching? The criticisms we are forced to make of Goodman's model considered as a possible description of behavior may very well not be appropriate when it is considered as a philosophical contribution.

Suppose that when two stimuli $s$ and $s'$ are presented the subject says either that they match (are the same) or that they do not and that he is
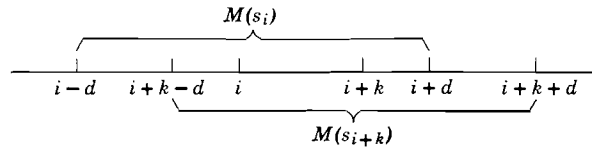
Fig. 11. Manors of stimuli that can be represented as points on a line.

consistent in his reports. Thus his behavior may be summarized as a matching relation $M$ over $\mathscr{S}$. In terms of a probability model, this is the same as supposing that all of the probabilities are either 0 or 1. By analogy to the probability models, we might suppose that there is a distance measure $d(s, s')$ [paralleling, e.g., $u(s) - u(s')$ in the Thurstone models] and a criterion $d$ such that the decision rule is

$$sMs' \quad \textit{if and only if} \quad d(s, s') \leqslant d. \tag{37}$$

Accepting the heuristic argument, it follows immediately that $M$ is a reflexive and symmetric relation but is not, in general, transitive. Goodman assumes that $M$ is reflexive and symmetric.

Now, suppose that instead of working with a sample of stimuli we have all possible stimuli of a given type, such as all 1000-cps tones of different intensities; suppose that the world is discrete in the sense that there is only a finite number of different stimuli, and suppose that they are ordered psychologically $s_1 < s_2 < \ldots < s_m$. Because these are supposed to be the only stimuli, the simplest possible distance measure is the number of intervening stimuli plus one: $d(s_i, s_j) = |j - i|$. If this is so and Eq. 37 is the decision rule, then there is a simple way to discover the distances between pairs of stimuli. We define the *manor* of stimulus $s$ to be the set of stimuli $M(s) = \{s' \mid sMs'\}$. For stimuli $s_i$ and $s_{i+k}$, where $s_iMs_{i+k}$, that is, where $k \leqslant d$, the manors are shown in Fig. 11. If $N(X)$ denotes the number of elements in a set $X$, then we see that

$$N[M(s_i) - M(s_{i+k})] = (i + k - d) - (i - d) = k,$$

and

$$N[M(s_{i+k}) - M(s_i)] = (i + k + d) - (i + d) = k,$$

hence

$$d(s_i, s_{i+k}) = k = \tfrac{1}{2}\{N[M(s_i) - M(s_{i+k})] + N[M(s_{i+k}) - M(s_i)]\}; \tag{38}$$

that is, the distance is one half the number of elements in the symmetric difference of the manors of the two stimuli.

Having the notion of distance given in Eq. 38, Goodman suggested that it can be used to reconstruct the ordering of the stimuli using the following three definitions:

1. *If s, s', and s" ∈ 𝒮 are all different and sMs', s'Ms", and sMs", then s' is between s and s" if and only if d(s, s") > d(s, s') and d(s, s") > d(s', s"), where d is defined by Eq. 38.*

2. *Stimulus s is said to be beside s' if and only if s ≠ s', sMs', and no s" exists between s and s'.*

3. *A finite set of stimuli over which a reflexive and symmetric relation M is defined form a linear array if two of the stimuli are each beside just one other stimulus and all others are beside exactly two others.*

It seems to us that for the empirical scientist this scheme to organize stimuli into linear arrays has a fatal flaw. The scheme rests crucially upon the assumption that one has all possible stimuli, for only in that case does Eq. ₋ᵠ give the underlying distance measure. When subsets of stimuli are used, as must be the case in practice, Eq. 38 does not define an invariant distance measure. Thus there may be arrays that are nonlinear by Goodman's definition which have subsets that are linear. An example is shown in Fig. 12. In the basic underlying two-dimensional grid each point matches itself and each adjacent point in its row and column. This array is clearly nonlinear, but the subset of darkened points and lines forms a linear array in Goodman's sense.

The results given below regarding Goodman's definition of a linear array are of interest. Suppose that by his method a set of stimuli form a linear array, ordered $s_1 \leqslant s_2 \leqslant \ldots \leqslant s_m$. Assume, in addition to
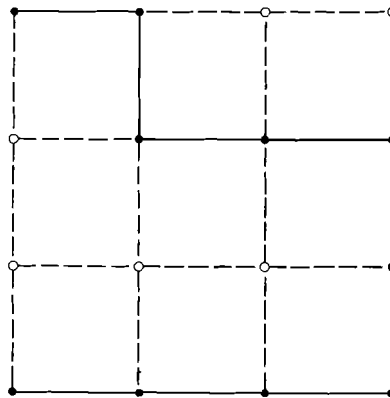


Fig. 12. An example of a nonlinear array having a subset which is linear.

reflexiveness and symmetry, that $M$ satisfies this condition which follows
from the basic decision rule:

**Weak Mapping Assumption.** If $s_i < s_j < s_k < s_l$ and $s_i M s_l$, then
$s_j M s_k$.

Suppose we define the distance measure $\delta(s_i, s_j) = |j - i|$. It should be
noted that in general this definition of distance differs from that given by
Eq. 38, which is the one used to organize the stimuli into a linear array.

Three concepts are now introduced. The first is, essentially, the decision
rule we have been assuming, but applied to the distance $\delta$.

1. *A linear array is* uniform *provided that a number $\delta$ can be found such
   that $sMs'$ if and only if $\delta(s, s') \leqslant \delta$.*
2. *A linear array is* regular *provided that when $sMs'$ and $\delta(t, t') \leqslant \delta(s, s')$
   then $tMt'$.*
3. *In a linear array $s$ is* just noticeably different *from $s'$, denoted $sJs'$,
   if not $sMs'$ and there exists an $s''$ such that $s''$ is beside $s$ and $s''Ms'$.*

Fine (1954) has proved the following theorem:

**Theorem 7.** *If a linear array satisfies the weak mapping assumption,
then the following are equivalent:* (1) *it is uniform*, (2) *it is regular, and*
(3) *$sJs'$ implies $s'Js$.*

In a later paper Fine and Harrop (1957) showed that it is always possible
to embed a linear array that satisfies the weak mapping assumption into
a uniform linear array. The motive for proving this is, evidently, to show
that we can always consider the set of stimuli actually used as a subset of
a set of stimuli for which the decision rule (Eq. 37) holds. It does not,
however, overcome the lack of invariance in Goodman's scheme for
arranging the stimuli into an array in the first place.

## 6. CONCLUSIONS

It came as a surprise to us when preparing this chapter, as it may have
to the reader, that in spite of a long history of experimental work on
discrimination there are relatively few data available that are suitable to
test existing mathematical theories. This is in striking contrast to the
abundance of detection and recognition data that have been collected
specifically to test and guide the recent flurry of theoretical research in
that area. It is evident that a program of empirical work needs to be
developed to parallel that now under way on detection.

Although there are currently three different response theories in the
detection studies, only two have been worked out for discrimination. We
are at present totally lacking a true threshold theory of discrimination.

As in the detection work, the most fully developed theories establish relations among response probabilities. Each of these theories defines numerical parameters, some of which are thought to be stimulus-determined and others, response biases. Adequate theories are still lacking to account for how these depend, respectively, upon the stimulus measures and upon payoffs, presentation probabilities, and the like. It is clear that much work needs to be done, especially on the way in which the response biases are determined. The possibility of treating the biases as the end product of a learning process, as was done with the choice and threshold models in detection, needs investigation.

## *References*

Abelson, R. M., & Bradley, R. A. A 2 × 2 factorial with paired comparisons, *Biometrics*, 1954, **10**, 487–502.

Boring, E. G. Urban's tables and the method of constant stimuli. *Amer. J. Psychol.*, 1917, **28**, 280–293.

Boring, E. G. The control of attitude in psychophysical experiments. *Psychol. Rev.*, 1920, **27**, 440–452.

Boring, E. G. Is there a generalized psychometric function? *Amer. J. Psychol.*, 1924, **35**, 75–78.

Boring, E. G. *A history of experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1950.

Boring, E. G., Langfeld, H. S., & Weld, H. P. *Foundations of psychology.* New York: Wiley, 1948.

Bradley, R. A. Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. *Biometrics*, 1954, **10**, 375–390.(a)

Bradley, R. A. Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, 1954, **41**, 502–537.(b)

Bradley, R. A. Rank analysis of incomplete block designs. III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, 1955, **42**, 450–470.

Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 1952, **39**, 324–345.

Bush, R. R., Luce, R. D., & Rose, R. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Vol. I. Stanford: Stanford Univer. Press, 1963, in press.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. acoust. Soc. Amer.*, 1957, **29**, 715–720.

Cobb, P. W. Weber's law and the Fechnerian muddle. *Psychol. Rev.*, 1932, **39**, 533–551.

Fechner, G. T. *Elemente der Psychophysik.* Leipzig: Breitkopf und Hartel, 1860.

Fine, N. J. Proof of a conjecture of Goodman. *J. Symbolic Logic*, 1954, **19**, 41–44.

Fine, N. J., & Harrop, R. Uniformization of linear arrays. *J. Symbolic Logic*, 1957, **22**, 130–140.

Ford, L. R., Jr. Solution of a ranking problem from binary comparisons. *Amer. Math. Mon.* Herbert Ellsworth Slaught Memorial Papers, 1957, 28–33.

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

Gerlach, Muriel W. *Interval measurement of subjective magnitudes with subliminal differences.* Stanford: Applied Mathematics and Statistics Laboratory, 1957, Technical Report 7.

Goodman, N. *The structure of appearance.* Cambridge: Harvard Univer. Press, 1951.

Green, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 335–369.

Guilford, J. P. A generalized psychophysical law. *Psychol. Rev.*, 1932, **39**, 73–85.

Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

Gulliksen, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, **53**, 199–213.

Gulliksen, H. A generalization of Thurstone's learning function. *Psychometrika*, 1953, **18**, 297–307.

Harris, W. P. A revised law of comparative judgment. *Psychometrika*, 1957, **22**, 189–198.

Holway, A. H., & Pratt, C. C. The Weber-ratio for intensive discrimination. *Psychol. Rev.*, 1936, **43**, 322–340.

Householder, A. S., & Young, G. Weber laws, the Weber law, and psychophysical analysis. *Psychometrika*, 1940, **5**, 183–193.

Hovland, C. I. A note on Guilford's generalized psychophysical law. *Psychol. Rev.*, 1938, **45**, 430–434.

James, W. *The principles of psychology.* New York: Holt, 1890.

Johnson, H. M. Did Fechner measure "introspectional" sensation? *Psychol. Rev.*, 1929, **36**, 257–284.

Johnson, H. M. Some properties of Fechner's intensity of sensation. *Psychol. Rev.*, 1930, **37**, 113–123.

Johnson, H. M. Are psychophysical problems genuine or spurious? *Amer. J. Psychol.*, 1945, **58**, 189–211.

Koenigs, M. G. Recherches sur les intégrales de certaines équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1884, **1**, Supplement S1-S41.

Koenigs, M. G. Nouvelles recherches sur les équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1885, **2**, 385–404.

Luce, R. D. Semiorders and a theory of utility discrimination. *Econometrica*, 1956, **24**, 178–191.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Luce, R. D. An observable property equivalent to a choice model for discrimination experiments. *Psychometrika*, 1962, **27**, 163–167.

Luce, R. D., & Edwards, W. The derivation of subjective scales from just noticeable differences. *Psychol. Rev.*, 1958, **65**, 222–237.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.*, 1947, **19**, 609–619.

Mosteller, F. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 1951, **16**, 3–11.(a)

Mosteller, F. Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. *Psychometrika*, 1951, **16**, 203–206.(b)

Mosteller, F. Remarks on the method of paired comparisons: III. A test of significance

for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 1951, **16**, 207–218.(c)

Müller, G. E. Über die Massbestimmung des Ortsinnes der Haut mittels der Methode der richtigen und falschen Fälle. *Arch. f. d. ges. Physiol.*, 1879, **19**, 191–235.

Scott, D., & Suppes, P. Foundational aspects of theories of measurement. *J. Symbolic Logic*, 1958, **23**, 113–128.

Shipley, Elizabeth F., & Luce, R. D. Discrimination among two- and three-element sets of weights. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Vol. I. Stanford: Stanford Univer. Press, 1963, in press.

Solomons, L. M. A new explanation of Weber's law. *Psychol. Rev.*, 1900, **7**, 234–240.

Thompson, G. H. A new point of view in the interpretation of threshold measurements in psychophysics. *Psychol. Rev.*, 1920, **27**, 300–307.

Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.(a)

Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389.(b)

Thurstone, L. L. The learning function. *J. gen Psychol.*, 1930, **3**, 469–493.

Thurstone, L. L. *The measurement of values.* Chicago: Univer. of Chicago Press, 1959.

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

Urban, F. M. On the method of just perceptible differences. *Psychol. Rev.*, 1907, **14**, 244–253.

Urban, F. M. The method of constant stimuli and its generalizations. *Psychol. Rev.*, 1910, **17**, 229–259.

van Laer, J. Response biases in psychophysical judgments: a test of Luce's model. Northwestern Univer., dittoed, undated.