Chapter 1

The Representational Measurement Approach to Psychophysical and Judgmental Problems

Geoffrey Iverson
R. Duncan Luce

1. INTRODUCTION

This chapter outlines some of the main applications to psychophysical and judgmental modeling of the research called the representational theory of measurement. Broadly, two general classes of models have been proposed in studying psychophysical and other similar judgmental processes: information processing models and phenomenological models. The former, currently perhaps the most popular type in cognitive psychology, attempt to describe in more or less detail the mental stages of information flow and processing; usually these descriptions are accompanied by flow diagrams as well as mathematical postulates. The latter, more phenomenological models attempt to summarize aspects of observable behavior in a reasonably compact fashion and to investigate properties that follow from the behavioral properties. Representational measurement theory is of the latter genre. We are, of course, identifying the end points of what is really a continuum of model types, and although we will stay mainly at the phenomenological end of the spectrum, some models we discuss certainly contain an element of information processing.

Two features of the behavior of people (and other mammals) need to be taken into account: (1) responses are variable in the sense that when a subject is confronted several times with exactly the same stimulus situation, he or
she may not respond consistently, and (2) stimuli typically have somewhat complex internal structures that influence behavior. Indeed, stimuli often vary along several factors that the experimenter can manipulate independently. Thus, behavioral and social scientists almost always must study responses to complex stimuli in the presence of randomness. Although somewhat overstated, the following is close to true: We can model response variability when the stimuli are only ordered and we can model "average" responses to stimuli having some degree of internal structure, but we really cannot model both aspects simultaneously in a fully satisfactory way. In practice we proceed either by minimizing the problems of structure and focusing on the randomness—as is common in statistics—or by ignoring the randomness and focusing on structure—as is done in the representational theory of measurement. Each is a facet of a common problem whose complete modeling seems well beyond our current understanding.

This chapter necessarily reflects this intellectual hiatus. Section II reports probabilistic modeling with stimuli that vary on only one dimension. Section III extends these probabilistic ideas to more complex stimuli, but the focus on structure remains a secondary concern. Sections IV through VII report results on measurement models of structure with, when possible, connections and parallels drawn to aspects of the probability models.

The domain of psychophysical modeling is familiar to psychologists, and its methods include both choice and judgment paradigms. Although not often acknowledged, studies of preferences among uncertain and risky alternatives—utility theories—are quite similar to psychophysical research in both experimental methods and modeling types. Both areas concern human judgments about subjective attributes of stimuli that can be varied almost continuously. Both use choices—usually among small sets of alternatives—as well as judgment procedures in which stimuli are evaluated against some other continuous variable (e.g., the evaluation of loudness in terms of numerals, as in magnitude estimation, and the evaluation of gambles in terms of monetary certainty equivalents).

Our coverage is limited in two ways. First, we do not attempt to describe the large statistical literature called psychometrics or, sometimes, psychological measurement or scaling. Both the types of data analyzed and models used in psychometrics are rather different from what we examine here. Second, there is another, small but important, area of psychological literature that focuses on how organisms allocate their time among several available alternatives (for fairly recent summaries, see Davison & McCarthy, 1988; Lowenstein & Elster, 1992). Typically these experiments have the experimenter-imposed property that the more time the subject attends to one alternative, the less is its objective rate of payoff. This is not a case of diminishing marginal utility on the part of the subject but of diminishing replenishment of resources. It is far more realistic for many situations than is
the kind of discrete choice/utility modeling we examine here. Space limitations and our hope that these models and experiments are covered elsewhere in this book have led us to omit them.

II. PROBABILISTIC MODELS FOR AN ORDERED ATTRIBUTE

Most experimental procedures used to investigate the ability of an observer to discriminate the members of a set of stimuli are variants of the following two:

1. Choice Paradigm. Here the observer is asked to choose, from an offered set of \( n \) stimuli, the one that is most preferred or the one that possesses the most of some perceived attribute shared by the stimuli.

2. Identification Paradigm. Here the observer is required to identify which of the \( n \) stimuli is presented.

The simplest case of each paradigm occurs when \( n = 2 \). The choice paradigm requires an observer to discriminate stimuli offered in pairs, whereas the identification paradigm forms the basis of yes-no detection. We focus on discrimination and detection in this section; see section III for the general case of each paradigm.

In this section we are concerned primarily with psychophysical applications in which stimuli are ordered along a single physical dimension such as line length, frequency, intensity, and so on. Accordingly we use positive real numbers \( x, y, s, n \), and so on to label stimuli.

For a discrimination task to be nontrivial, the members of an offered pair \( x, y \) of stimuli must be close in magnitude. An observer's ability to decide that \( x \) is longer, or louder, or of higher pitch, and so on than \( y \) is then difficult, and over repeated presentations of \( x \) and \( y \), judgments show inconsistency. The basic discrimination data are thus response probabilities \( P_{x,y} \), the probability that \( x \) is judged to possess more of the given attribute than \( y \).

For a fixed standard \( y \), a plot of \( P_{x,y} \) against \( x \) produces a classical psychometric function. A typical psychometric function, estimated empirically, is shown in Figure 1. The data were collected by the method of constant stimuli: subjects were presented a standard line of length 63 mm together with one of five comparison lines that varied in length from 62 to 65 mm. Note that, in these data, the point of subjective equality is different from (here it is smaller than) the 63 mm standard length—not an uncommon feature of this psychophysical method.

The family of psychometric functions generated by varying the standard affords one way to study the response probabilities \( P_{x,y} \). It is often more revealing, however, to study \( P_{x,y} \) as a family of isoclines, that is, curves on which \( P_{x,y} \) is constant. Fixing the response probability at some value \( \tau \) and trading off \( x \) and \( y \) so as to maintain this fixed response probability generates
FIGURE 1  A typical psychometric function. The proportion of "larger" judgments is plotted as a function of a physical measure of the comparison stimulus. In this case the standard was a 63 cm length. From Figure 6.2 of *Elements of Psychophysical Theory*, by J.-C. Falmagne, New York: Oxford University Press, 1985, p. 151; redrawn from the original source, Figure 2.5 of "Psychophysics: Discrimination and Detection," by T. Egen, which appeared as chapter 2 of *Experimental Psychology*, by J. W. Kling and L. A. Riggs, New York: Holt, Rinehart and Winston, 1971. Reprinted with permission.

a sensitivity function \( \xi_\pi : P_{x,y} = \pi \) if and only if \( \xi_\pi(y) = x \). Writing \( \xi_\pi(x) = x + \Delta_\pi(x) \) defines the Weber function, or \( \pi \)-jnd of classical psychophysics. For instance, the symbol \( \Delta_{0.75}(x) \) denotes the increment added to a stimulus \( x \) so as to render the sum detectable from \( x \) 75% of the time; in classical psychophysics the arbitrary choice \( \pi = 0.75 \) served to define the just-noticeable difference.

Trade-off functions offer a convenient way to organize and study a wide range of psychophysical data. Equal loudness contours, intensity-duration trading relations, speed-accuracy trade-offs are but three of many examples that could be mentioned. For the detection task, the fundamental trade-off involves two kinds of error: one, an error of omission, arises when an observer fails to recognize the presence of a signal as such; the other, an error of commission, arises when an observer incorrectly reports the presence of the signal. The trade-off between these two sorts of error underlies the receiver operating characteristic (ROC), the basic object of study in yes-no detection.

In this section we provide an overview of models describing families of psychometric functions, sensitivity functions, ROCs, and other trade-offs.
such as speed-accuracy. Link (1992) is a modern text that covers much of the same material in detail, albeit from a different point of view.

A. Discrimination

1. Fechner’s Problem

The simplest class of models for response probabilities involves an idea proposed by Fechner (1860/1966); namely, that a comparison of stimuli \( x, y \) is based on the difference \( u(x) - u(y) \) of internal “sensations” evoked by \( x \) and \( y \). Here the numerical scale \( u \) is assumed to be a strictly increasing function of the physical variable. Confusion between \( x, y \) arises because the difference \( u(x) - u(y) \) is subject to random error (which Fechner took to be normally distributed\(^1\)). In other words,

\[
P_{x,y} = \text{Prob}[u(x) - u(y) + \text{random error} \geq 0].
\]

In terms of the distribution function \( F \) of the error, this amounts to

\[
P_{x,y} = F[u(x) - u(y)],
\]

where each of the functions \( F \) and \( u \) is strictly increasing on its respective domain. The form for the response probabilities given in Eq. (1) is called a Fechner representation for those probabilities. Fechner’s problem (Falmagne, 1985) is to decide, for a given system of response probabilities, if a Fechner representation is appropriate and, if so, to determine how unique the representation is.

These and other related matters have received much attention in the theoretical literature (Falmagne, 1985; Krantz, Luce, Suppes, & Tversky, 1971; Levine, 1971, 1972; Suppes, Krantz, Luce, & Tversky, 1989). A key observable property enjoyed by all systems of response probabilities conforming to the Fechnerian form of Eq. (1) is known as the quadruple condition (Marschak, 1960): for all \( x, y, x', y' \),

\[
\text{if } P_{x,y} \geq P_{x',y}, \text{ then } P_{x,y'} \geq P_{y,y'}. \tag{2}
\]

This property is easily seen to be necessary for a Fechner representation. In terms of scale differences, the left-hand inequality of Eq. (2) asserts that \( u(x) - u(y) \geq u(x') - u(y') \), which rearranges to read \( u(x) - u(x') \geq u(y) - u(y') \) and in turn that inequality implies the right-hand inequality of Eq. (2). It is a far more remarkable fact that the quadruple condition is, in the presence of natural side conditions, also sufficient for a Fechner representation; see Falmagne (1985) for a precise statement and proof of this fact.

The uniqueness part of Fechner’s problem is readily resolved: the scale \( u \) is unique up to positive linear transformations, that is, \( u \) is an interval scale (see section IVB.3). This is not surprising in view of the fact that positive scale differences behave like lengths and can be added:

\(^1\) Also called Gaussian distributed.
See section IV.C.6 for an account of the theory of measurement of length and other extensive attributes.

Although the quadruple condition presents an elegant solution to Fechner’s problem, it is not easily tested on fallible data; for a general approach to testing order restrictions on empirical frequencies, see Iverson and Falmagne (1985).

2. Weber Functions

Another approach to Fechner’s problem is afforded by the study of the sensitivity functions \( \xi_n \) (or equivalently the Weber functions \( \Delta_n \)). To assume the validity of the representation Eq. (1) is equivalent to assuming the following representation for sensitivity functions:

\[
\xi_n(x) = u^{-1}[u(x) + g(\pi)];
\]

where \( g = F^{-1} \). In these terms an alternative formulation of Fechner’s problem can be framed as follows: What properties of sensitivity functions guarantee a representation of the form Eq. (3) for these functions? A condition that is clearly necessary is that two distinct sensitivity functions cannot intersect—sensitivity functions are ordered by the index \( \pi \). Moreover, sensitivity functions of the desired form can be concatenated by the ordinary composition of functions, and this concatenation is commutative (i.e., the order of the composition makes no difference):

\[
\xi_n[\xi_m(x)] = u^{-1}[u(x) + g(\pi) + g(\pi')] = \xi_m[\xi_n(x)].
\]

These two properties allow the collection of sensitivity functions to be recognized as an ordered abelian group, and the machinery of extensive measurement applies (see section IV.C). For a detailed discussion, see Krantz et al. (1971), Suppes et al. (1989), Levine (1971, 1972), and Falmagne (1985). Kuczma (1968) discussed the problem from the viewpoint of iterative functional equations.

The previous remarks reflect modern ideas and technology. Fechner studied the functional equation

\[
u(x + \Delta(x)) = u(x) = 1,
\]

A mathematical group \( G \) is a set of objects (here functions) together with a binary operation \( * \) (here composition of functions), which is associative: for all objects \( x, y, z \) in \( G \), \( x(yz) = (xz)y \). There is an identity element \( e \) (here the identity function) and each element \( x \) in \( G \) possesses an inverse \( x^{-1} \) such that \( x * x^{-1} = x^{-1} * x = e \). The group is abelian when \( * \) is commutative, i.e., \( xy = yx \).
FIGURE 2  Weber functions for loudness of pure tones in which the logarithm of the Weber fraction is plotted against the sound pressure level in decibels above threshold for eight frequencies from 200 to 8000 Hz. In such a plot, Weber's law would appear as a horizontal line. From Figure 1 of "Intensity Discrimination as a Function of Frequency and Sensation Level," by W. Jesteadt, C. C. Wier, and D. M. Green, 1977, Journal of the Acoustical Society of America, 61, p. 171. Reprinted with permission.

called Abel's equation, and incorrectly reasoned that it could be replaced by a differential equation (Luce & Edwards, 1958). He did, however, correctly perceive that Weber's law—namely, the assertion that $\Delta N(x)$ is proportional to $x$ for any value of $\pi$—provides a rapid solution to Fechner's problem. In our notation Weber's law is equivalent to the assertion

$$P_{x,y} = P_{x,y}$$

for any positive real number $c$ and all $x, y$. It follows at once that $P_{x,y}$ depends only on the ratio $x/y$ of physical measures and that the scale $u(x)$ is logarithmic in form. Although Weber's law remains a source of useful intuition in psychophysics, it provides at best an approximation to the empirical data. For example, in psychoacoustics, pure tone intensity discrimination exhibits the "near-miss" to Weber's law (Figure 2). On the other hand, Weber's law holds up remarkably well for intensity discrimination.
FIGURE 3 Weber function for loudness of white noise in which the Weber fraction is presented in decibel terms versus the sound pressure level in decibels relative to threshold. Again, in such a plot, Weber's law would appear as a horizontal line, which is true for most (recall, this is a logarithmic scale) of the stimulus range. From Figure 5 of "Discrimination," by R. D. Luce and E. Galanter, in R. D. Luce, R. R. Bush, and E. Galanter, Handbook of Mathematical Psychology (Vol. 1), New York: John Wiley & Sons, 1963, p. 203. Reprinted with permission.

of broadband noise (Figure 3). For further remarks on Weber's law and the near-miss, see section VII.B.1.

3. Random Variable Models

Suppose that a stimulus $x$ elicits an internal representation as a random variable $U_x$. In these terms, the response probabilities can be written

$$P_{X,Y} = \text{Prob}(U_x \geq U_y).$$

Such a representation was first proposed and studied in the literature on individual choice, where the term random utility model has become standard (Block & Marschak, 1960; Luce & Suppes, 1965; Marschak, 1960). Although this representation does impose constraints on the response probabilities, for example, the triangle condition $P_{x,y} + P_{y,z} + P_{z,x} \geq 1$, it is not well understood (see Marley, 1990, for a review) and is clearly very weak. For these reasons it is useful to explore the consequences of specific distributional assumptions on the random variables involved in Eq. (4).

We follow the convention of using uppercase bold letters such as $X$, $Y$, and $Z$ to denote random variables. We shall write vectors as bold lowercase letters such as $x$, $y$, and $z$. A vector-valued random variable is not distinguished notationally but rather by context.
In a trio of seminal papers, Thurstone (1927a, b, c) made the assumption that $U_x, U_y$ are jointly normal. Doing so gives rise to a relation known as Thurstone’s law of comparative judgment; see Eq. (5) for a special, but important case. In many circumstances it is reasonable to suppose that $U_x$ and $U_y$ are not only normal but independent. The stability of the normal family—the fact that the sum (or difference) of two independent normally distributed random variables remains normally distributed—allows Eq. (4) to be developed in terms of the means $\mu(x), \mu(y)$ and variances $\sigma^2(x), \sigma^2(y)$ of $U_x$ and $U_y$:

$$P_{x,y} = \Phi[(\mu(x) - \mu(y))/\sqrt{\sigma^2(x) + \sigma^2(y)}],$$

where $\Phi$ is the distribution function of the unit normal (mean zero, variance unity). This representation is Case III in Thurstone’s classification.

When $\sigma(x)$ is constant across stimuli, one obtains the simple Case V representation:

$$P_{x,y} = \Phi[\mu(x) - \mu(y)],$$

where $\mu(x) = \mu(x)/\sqrt{2}\sigma$. This model is a special case of a Fechnerian representation; compare it with Eq. (1).

Thurstone offered little to justify the assumption of normality; indeed, he admitted it might well be wrong. However, in many stochastic process models, information about a stimulus arises as a sum of numerous independent contributions. Such sums are subject to the central limit theorem, which asserts that their limiting distribution is normal; an explicit example of this sort of model is discussed in section II.C.

Other authors, for example, Thompson and Singh (1967) and Pelli (1985), have proposed models in which discriminative information is packaged not as a sum but as an extreme value statistic. Invoking a well-known limit law for maxima (Galambos, 1978/1987) leads to a model in which the random variables $U_x$ and $U_y$ of Eq. (4) are independent, double-exponential variates with means $\mu(x), \mu(y)$. The following expression for the response probabilities results:

$$P_{x,y} = \frac{\exp[\nu(x)]}{[1 + \exp(\mu(y) - \mu(x))] = \nu(x)/[\nu(x) + \nu(y)],$$

where $\nu(x) = \exp[\mu(x)]$. The expression given in Eq. (7) is often called a Bradley-Terry-Luce representation for the response probabilities.

The expression in Eq. (7) also arises in choice theory (section III.A.2), but is based on quite different considerations (Luce, 1959a). The following product rule is a binary property that derives from the more general choice theory: for any choice objects $a, b, c$,

$$\frac{P_{a,b}}{P_{b,a}} \cdot \frac{P_{b,c}}{P_{c,b}} = 1.$$

The product rule in Eq. (8) is equivalent to the representation in Eq. (7).
B. Detection

1. Receiver Operating Characteristics

The basic detection task requires an observer to detect the presence or absence of a signal embedded in noise. On some trials the signal accompanies the noise; on other trials noise alone is presented. On each trial the observer makes one of two responses: signal present, "yes," or signal not present, "no." Two kinds of errors can be made in this task. One, called a miss occurs when a no response is made on a signal trial; the other, called a false alarm occurs when a yes response is made on a noise-alone trial. Corresponding correct responses are called hits (yes responses on signal trials) and correct rejections (no responses on noise trials).

Because hits and misses are complementary events, as are correct rejections and false alarms, the yes-no task involves only two independent response rates and it is conventional to study the pair of conditional probabilities $P_H = \Pr(\text{yes|signal})$ and $P_{FA} = \Pr(\text{yes|noise})$. The two probabilities move together as a function of an observer's tendency to respond yes. Such biases can be brought under experimental control by employing an explicit payoff schedule: punishing false alarms depresses the frequency of yes responses, and $P_H$ and $P_{FA}$ each decrease; rewarding hits has the opposite effect. By varying the payoff structure, the pair $(P_H, P_{FA})$ traces out a monotonically increasing curve from $(0, 0)$ to $(1, 1)$ in the unit square. Such a curve is called a receiver operating characteristic, abbreviated ROC. An alternative method of generating an ROC involves varying the probability of a signal trial. Figure 4 shows typical ROCs generated by both methods.

A single ROC is characterized by fixed signal and noise parameters; only an observer's bias changes along the curve. By varying signal strength, a family of ROCs is obtained, as in Figure 5. For a wealth of information on detection tasks and the data they provide consult Green and Swets (1966/1974/1988) and Macmillan and Creelman (1991).

2. Psychometric Functions

In classical psychophysics, it was common practice to study the psychometric function obtained by measuring the hit rate $P_H$ as signal intensity was varied. Instructions were intended to practically forbid the occurrence of false alarms. This strategy is fraught with difficulties of estimation: $P_H$ must be estimated on the most rapidly rising part of an ROC, so that small errors in $P_{FA}$ become magnified in the determination of $P_H$.

3. Statistical Decision Making

Detection can be modeled as a problem of statistical decision making. In this view, evidence for the signal is represented as a random variable whose
values are distributed on a one-dimensional "evidence" axis. On signal trials, the evidence for the signal is a value of a random variable $U_j$; on noise trials, evidence is a value of a random variable $U_n$. Large values of evidence arise more frequently on signal trials and thus favor the presence of the signal. An observer selects a criterion value $\beta$ on the evidence axis, which is sensitive to payoff structure and signal probability, such that whenever the evidence sampled on a trial exceeds $\beta$, the observer responds yes, indicating a belief that the signal was presented.

Of the various candidates for the evidence axis, one deserves special mention. According to the Neyman-Pearson lemma of statistical decision
FIGURE 5  ROCs obtained using a five-point rating scale and varying signal strength over seven levels (the weakest and strongest levels are omitted in the plot). The stimuli were 60 Hz vibrations to the fingertip, the curves are identified by the amplitude of the stimulus in microns. The procedure involved two conditions, represented by the open and closed symbols. In each, the probability of no signal was 0.33 and of a signal, 0.67. In the case of the open symbols, the three weaker signals were more likely than the four stronger ones (signal probabilities of 0.158 and 0.066, respectively), whereas for the closed symbols the three stronger signals were more likely than the four weaker ones (again, 0.158 and 0.066). Thus, there was a single false alarm estimate for all seven intensities corresponding to each of rating levels. From Figure 3.24 of Psychophysics: Methods and Theory, by G. A. Gescheider, Hilldale, NJ: Erlbaum, 1976, p. 79. Redrawn from the original source “Detection of Vibrotactile Signals Differing in Probability of Occurrence,” G. A. Gescheider, J. H. Wright, and J. W. Polak, 1971, The Journal of Psychology, 78, Figure 3, p. 259. Reprinted with permission.

theory, the optimal way to package evidence concerning the signal is to use the likelihood ratio—the ratio of the density of sensory data assuming a signal trial to the density of the same data assuming a noise-alone trial. Large values of the likelihood ratio favor the presence of the signal.

However there remains considerable flexibility in the choice of a decision statistic: any strictly increasing function of the likelihood ratio produces an equivalent decision rule and leads to identical detection performance. A common choice of such a transformation is the logarithm, so that evidence can take on any real value. It is worthy of note that ROCs that are concave (as are those of Figures 4 and 5) are compatible with the use of the likelihood-ratio as a decision statistic (cf. Falmagne, 1985).

On the other hand, there is little reason to suppose human observers can
behave as ideal observers, except in the simplest of circumstances (see Green & Swets, 1966/1974/1988, for further discussion). More likely than not, human observers use simple, easy-to-compute decision statistics that will not, in general, be monotonically related to the likelihood ratio (see, e.g., section II.C).

4. Distributional Assumptions and \(d'\)

It should be noted that the representation of an ROC in terms of decision variables \(U_s, U_a\), namely,

\[
P_{H} = \text{Prob}(U_s > \beta), \quad P_{FA} = \text{Prob}(U_a > \beta)
\]

is not at all constraining, despite the rather heavy background imposed by statistical decision theory. If one chooses \(U_a\) to possess a strictly increasing, but otherwise arbitrary distribution function, it is always possible to find a random variable \(U_s\) such that a given ROC is represented in the form of Eq. (9) (cf. Iverson & Sheu, 1992).

On the other hand, empirical families of ROCs obtained by varying some aspect of the signal (such as intensity or duration) often take on a simple, visually compelling form. The ROCs given in Figure 5 are, above all, clearly ordered by varying stimulus amplitude. This suggests, at least in such examples, that ROCs are isoclines of some function monotonically related to signal strength; moreover, because these isoclines do not intersect, a Fechnerian representation may hold (recall the discussion in Section II.A):

\[
\text{Stimulus strength} = F[u(P_{H}) - u(P_{FA})].
\]

In other words, there exists the possibility of transforming an ROC into a line of unit slope by adopting \(u(P_{H}), u(P_{FA})\) as new coordinates.

It is not difficult to show that this possibility does occur if \(U_s, U_a\) are members of a location family of random variables, differing only in their mean values. Based on explicit examples and, above all else, on simplicity, it is commonly assumed that \(U_s, U_a\) are normally distributed, with a common variance. This assumption is responsible for the custom of plotting ROCs on double-probability paper (with inverse normals along the axes). If the normal assumption is correct, ROCs plot as parallel lines of unit slope, with intercepts

\[
d' = z_{H} - z_{FA} = [\mu(s) - \mu(a)]/\sigma.
\]

* Suppose the value of a real function \(F\) of two real variable is fixed: \(F(x,y) = \text{constant}\). Then such pairs \((x,y)\) trace out a curve called an isocline or level curve of \(F\). Different isoclines correspond to different values of the function \(F\).
where $z = \Phi^{-1}$ (probability) and $\Phi$ is the distribution function of the unit normal. The measure $d'$ depends only on stimulus parameters and is thus a measure of detectability uncontaminated by subjective biases.

The remarkable fact is that when empirical ROCs are plotted in this way, they do more or less fall on straight lines, though often their slopes are different from unity. This empirical fact can be accommodated by retaining the normality assumption but dropping the constant variance assumption. Using the coordinate transformation $z = \Phi^{-1}$ (probability)—recall $\Phi$ is the distribution function of the unit normal—the following prediction emerges:

$$\sigma(z)z_{\delta_{1}} - \sigma(n)z_{\delta_{2}} = \mu(z) - \mu(n),$$

which is the equation of a line of slope $\sigma(n)/\sigma(z)$ and intercept $[\mu(z) - \mu(n)]/\sigma(z)$. Unlike the case discussed earlier for which $\sigma(z) = \sigma(n)$, that is, Eq. (10), there is now some freedom in defining an index of detectability, and different authors emphasize different measures. Those most commonly employed are the following three:

- $[\mu(z) - \mu(n)]/\sigma(z)$,
- $[\mu(z) - \mu(n)]/\sigma(n)$,
- $\mu(z) - \mu(n) + \sqrt{\sigma^{2}(z) + \sigma^{2}(n)}$.

Note that the latter index, the perpendicular distance to the line (Eq. 11) from the origin, is closely related to performance in a discrimination (two alternative/interval forced-choice) paradigm using the same signal and noise sources as employed in the detection task. Formally, the prediction for the forced-choice paradigm is given by Eq. (5) with $x,y$ replacing $x,y$, respectively. One obtains

$$z_{c} = \Phi^{-1}(\frac{1}{\sqrt{\sigma^{2}(z) + \sigma^{2}(n)}}),$$

where $z_{c}$ is the transformed probability of a correct response in the two alternative tasks. The ability to tie together the results of different experimental procedures is an important feature of signal detection theory, one that has been exploited in many empirical studies. For additional results of this type, see Noreen (1981) and Macmillan and Creelman (1991), who confine their developments to the constant variance assumption, and Iverson and Sheu (1992), who do not. In section II.C we sketch a theory that unites detection performance and speed-accuracy trade-off behavior under a single umbrella.

5. Sources of Variability

A question first raised by Durlach and Braida (1969), Gravetter and Lockhead (1973), and Wickelgren (1968) concerns the locus of variability in this class of signal detection models. Eq. (9) is written as if all of the variability lies in the representation of the stimuli, and the response criterion $\beta$ is
treated as a deterministic numerical variable. For the case of location families of random variables, the data would be fit equally well if all the variability were attributed to $\beta$ and none to the stimuli. Indeed, because variances of independent random variables add, any partition between stimulus variability and criterion variability is consistent with both yes-no and forced-choice data. The problem, then, is to design a method that can be used to estimate the partition that actually exists.

Perhaps Nosofsky (1983) provided the cleanest answer. His idea was to repeat the stimulus presentation $N$ times with independent samples of noise and have the subject respond to the entire ensemble. If subjects average the $N$ independent observations, the mean is unaffected but the variance decreases as $\sigma^2(s)/N$. On the other hand, there is no reason why the criterion variance $\sigma^2(\beta)$ should vary with $N$. Substituting into Eq. (10), we obtain

$$\frac{1}{d_{N}^2} = \frac{\sigma^2(\beta) + \sigma^2(s)/N}{\mu(s) - \mu(\alpha)},$$

which represents a linear trade-off between the variables $1/(d_{N}^2)$ and $1/N$.

Nosofsky carried out an auditory intensity experiment in which four signals were to be identified. The two middle signals were always at the same separation, but the end signals were differently spaced leading to a wide and a narrow condition. The quantity $d'_{N}$ was computed for the pair of middle stimuli of fixed separation. Figure 6 shows $1/(d_{N}^2)$ versus $1/N$ for both the wide and narrow conditions. The predicted linearity was confirmed. The value (slope/intercept)$^{1/2}$, which estimates $\sigma(s)/\sigma(\beta)$, is 3.96 and 3.14 in the wide and narrow conditions, respectively; the ratio $\sigma(s, \text{wide})/\sigma(s, \text{narrow})$ is 7.86, and that of $\sigma(\beta, \text{wide})/\sigma(\beta, \text{narrow})$ is 6.23. Thus it appears that the standard deviations for stimulus and criterion partition about 3 or 4 to 1. Nosofsky also reanalyzed data of Ulehla, Halpern, and Cerf (1968) in which subjects identified two tilt positions of a Line; again the model fit well. The latter authors varied signal duration and that manipulation yielded estimates of $\sigma(s)/\sigma(\beta)$ of 14.22 in the shorter duration and 4.57 in the longer one. The criterion variance was little changed across the two conditions.

6. COSS Analysis

A far-reaching generalization of Nosofsky’s ideas was recently proposed by Berg (1989) under the name of COSS analysis (conditional on a single stimulus). Rather than assume an observer gives equal weight to all sources of information relevant to detecting a signal, Berg’s theory calls for a system of differential weights. COSS analysis provides an algorithm for estimating these weights in empirical data. There is a growing body of evidence that
observer's do not usually employ equal weights, even when, as in Nosofsky's paradigm, they should; rather, the pattern of weights takes on a variety of shapes depending on the structure of stimuli and the demands of a particular task. Berg (1989), Berg (1990), and Berg and Green (1990) discuss tasks that produce rather different weight patterns.

Since its inception about eight years ago, COSS analysis has had a major impact in psychoacoustics, where it was first applied. However, the technique is very flexible and one can expect it will find application to any task calling for the detection of complex stimuli that vary on many dimensions.

C. Stochastic Process Models

The models we have considered thus far are largely phenomenological. They allow for useful interpretations of data, but they do not attempt to capture the complexity of stimulus encoding as revealed by physiological studies. Yet efforts to create more realistic models of information transmission, however crude and incomplete, seem to be of considerable merit. We now sketch the results of one such enterprise.

Physiological studies conducted in the 1960s and 1970s (summarized in Luce, 1986, 1993) of the temporal coding of simple tones by individual fibers of the eighth nerve revealed that histograms of interpulse times were roughly exponential in their gross shape. (This rough exponential shape ignores fine structure: There is refractoriness, and the actual distribution is spiky, with successive peaks displaced at intervals of $1/T$, $T$ being the period of the input tone.) Assuming independence of times between successive pulses, such exponential histograms suggest that the encoding of simple auditory stimuli can be modeled as Poisson processes of neural pulses, with rates determined by stimulus intensity (see Green & Luce, 1973); however, more recent work casts doubt on the independence of successive pulse durations (Lowen & Teich, 1992).

A Poisson process can be thought of as a succession of points on a line, the intervals between any two consecutive points being distributed independently and exponentially. The reciprocal of the mean interval between successive events defines the rate parameter of the process.

FIGURE 6  Plot of estimated $1/d^2$ versus $1/N$, where $N$ is the number of independent repetitions of a pure tone that was to be absolutely identified from one of four possible intensities. The middle two stimuli had the same separation in both conditions, which were determined by the separation—wide or narrow—of the two end stimuli. $d'$ was calculated for the two middle stimuli for each of eight subjects and then averaged; an average of 187.5 observations underlie each point of the wide condition and an average of 150 for the narrow one. The least-squares fits are shown. From Figure 2 of "Information Integration of the Identification of Stimulus Noise and Criteria Noise in Absolute Judgment," by R. M. Nosofsky, 1983, Journal of Experimental Psychology: Human Perception and Performance, 9, p. 305. Copyright 1983 by the American Psychological Association. Reprinted by permission.
A Poisson process allows two basic ways for estimating the rate parameter:

1. Count the number of events in a fixed time interval (the counting strategy).
2. Compute the reciprocal of the mean interarrival time between pulses (the timing strategy).

In a simple detection task involving pure tones in noise, Green and Luce (1973) argued that if an observer can use these two decision strategies, then it should be possible to induce the observer to switch from one to the other. An observer whose brain counts pulses over a fixed time interval may be expected to perform differently from one who whose brain calculates the (random) time required to achieve a fixed number of events. Indeed, the counting strategy predicts that ROCs will plot as (approximate) straight lines in Gaussian coordinates with slopes $\sigma(n)/\sigma(s)$ less than unity, whereas for the timing strategy the ROCs are again predicted to be (approximately) linear on double-probability paper but with slopes exceeding 1. Green and Luce found that observers could be induced to switch by imposing different deadline conditions on the basic detection task: when observers were faced with deadlines on both signal and noise trials, they manifested counting behavior (see Figure 7, top); when the deadline was imposed only on signal trials, observers switched to the timing strategy (see Figure 7, bottom).

The very nature of these tasks calls for the collection of response times. Green and Luce developed response time predictions for the two types of strategy. Predictions for the counting strategy are trivial because such observers initiate a motor response after the fixed counting period: mean latencies should thus show no dependence on stimulus or response, in agreement with observation. For the timing strategy, however, different speed-accuracy trade-offs are predicted on signal trials and on noise trials. Again, the data bore out these predictions (see Figure 11 of Green and Luce, 1973). The issue of averaging information versus extreme values was also studied in vision; see Wandell and Luce (1978).

III. CHOICE AND IDENTIFICATION

A. Random Utility Models

1. General Theory

A participant in a choice experiment is asked to select the most preferred alternative from an offered set of options. Such choice situations are commonly encountered in everyday life: selecting an automobile from the host of makes and models available, choosing a school or a house, and so on. To account for the uncertainties of the choice process, which translate into data
inconsistencies, choice models are typically framed in terms of choice probabilities $P_{s,A}$, the probability of selecting an option $a$ from Set $A$ of alternatives.

A random utility model for the choice probabilities involves the assumption that each alternative $a$ is associated with a random variable $U_a$ that measures the (uncertain) value or utility of that alternative. In these terms it is natural to assume that

$$P_{s,A} = \text{Prob}(U_a \geq U_b, \text{ all } b \in A),$$

(14)

generalizing the binary choice situation discussed earlier in section II.A.3 [see Eq. (4)].
Without specific assumptions on the family of random variables \( \{U_a\} \) appearing in a random utility representation—for example, that they are independent or that their joint distribution is known up to the values of parameters—it would appear that Eq. (14) does little to constrain observed choice probabilities. However, following Block and Marschak (1960), consider the following chain of expressions involving linear combinations of choice probabilities:

\[
P_{a,A^*} = P_{a,A^*} - (P_{a,A^*} - P_{a,A}) + (P_{a,A} - P_{a,A^*}) - (P_{a,A} - P_{a,A}) + (P_{a,A} - P_{a,A}) + (P_{a,A} - P_{a,A}) + \ldots
\]

where \( A = \{a, b, c, d, \ldots\} \) and where the notation \( A - B \), \( B \) a subset of \( A \), represents the set of members of \( A \) that are not also members of \( B \).

It can be shown that Eq. (14) requires each of these so-called Block-Marschak functions to be nonnegative. In other words, the nonnegativity of Block-Marschak functions is a necessary condition for the existence of a random utility representation of choice probabilities. A remarkable result of Falmagne (1978) shows the same condition to be sufficient for a random utility representation.

A random utility representation of choice probabilities is far from unique: Any strictly increasing function applied to the random variables \( \{U_a\} \) provides another, equivalent, random representation of the same choice probabilities [see Eq. (14)]. To address this lack of uniqueness, consider a variant of the choice paradigm in which the task is to rank order the "alternatives from most preferred to least preferred." Define the random variable \( U_a^* = k \) if alternative \( a \) is assigned rank \( k \), \( k = 1, 2, \ldots \). Following the earlier work of Block and Marschak (1960), Falmagne established three results:

1. The random variables \( \{U_a^* \} \) provide a random utility representation (whenever one exists).
2. All random utility representations for a given system of choice probabilities yield identical ranking variables \( U_a^* \).
3. The joint distribution of the ranking variables can be constructed from the choice probabilities.

For a detailed discussion of these facts, see Falmagne (1978).

Recently Regenwetter (1996) generalized the concept of a random utility representation to \( m \)-ary relations. The applications of his theory include a model of approval voting and an analysis of political ranking data.

Despite this impressive theoretical analysis of Eq. (14), very little in the
way of empirical application has been attempted; Iverson and Bamber (1997) discuss the matter in the context of signal detection theory, where the random variables appearing in Eq. (14) can be assumed independent. Rather, the impact of specific distributional and other assumptions on Eq. (14) has dominated the field.

2. Luce's Choice Model

The assumption that the random variables $U_a$ appearing in Eq. (14) are jointly normal (following Thurstone, see section II.A.3), does not lend itself to tractable analysis, except in special cases such as pair-comparison tasks. This circumstance arises from the fact that the maximum of two or more normal random variables is no longer normally distributed. Only three families of distributions are "closed" under the operation of taking maxima, and of these the double-exponential family is the most attractive. We mentioned in section II.A.3 that the assumption of double-exponentially distributed random variables mediating discrimination of two stimuli leads to the Bradley-Terry-Luce model for pair-comparison data [see Eq. (7)].

If one assumes that the random utilities in Eq. (14) are members of a location family, that is, of the form $u(a) + U$, $u(b) + U'$, $u(c) + U''$, \ldots where $U, U', U''$ are independent with a common double-exponential distribution, namely,

$$\text{Prob}(U \leq t) = \exp(-e^{-t})$$

for all $t$, it follows from Eq. (14) that, for a Set $A$ of alternatives $a, b, c, \ldots$,

$$P_{a,A} = \frac{\nu(a)}{\sum_{b \in A} \nu(b)}.$$  \hspace{1cm} (16)

This expression also arises from Luce's (1959a) theory of choice.

Yellott (1977) has given an interesting characterization of the double-exponential distribution within the context of Eq. (14). He considered all choice models involving random utilities of an unspecified location family and he inquired as to the effect on choice probabilities of uniformly expanding the choice set by replicating each alternative some fixed but arbitrary number of times. Thus, for example, if a choice set comprises a glass of milk, a cup of tea, and a cup of coffee, a uniform expression of that set would contain $k$ glasses of milk, $k$ cups of tea, and $k$ cups of coffee for some integer $k \geq 2$. Yellott showed that if choice probabilities satisfying Eq. (14) were unchanged by any uniform expansion of the choice set, then, given that the (independent) random utilities were all of the form $u(a) + U$, the distribution of the random variable $U$ is determined. It must be double-exponential.
3. Elimination Models

The choice model Eq. (16) has been the subject of various criticisms on the basis of which new theories have been proposed. Suppose, for example, that one is indifferent when it comes to choosing between a cup of tea and a cup of coffee. Intuitively, the addition of a further cup of tea should not affect the odds of choosing coffee. Yet the model Eq. (16) predicts that the probability of choosing coffee drops to one third unless, of course, equivalent alternatives are collapsed into a single equivalence class.

Tversky (1972a, b) offered a generalization of Luce’s choice model that escapes this and other criticisms. In his theory of choice by elimination, each choice object is regarded as a set of features or aspects to which weights are attached. The choice of an alternative is determined by an elimination process in which an aspect is selected with a probability proportional to its weight. All alternatives not possessing the chosen aspect are eliminated from further consideration. The remaining alternatives are subject to the same elimination process until a single alternative remains. This theory reduces to Luce’s choice model in the very special case in which alternatives do not share common aspects.

Practical implementation of the elimination-by-aspects model is made difficult by the large number of unknown parameters it involves. This difficulty is alleviated by imposing additional structure on the alternatives. Tversky and Sattath (1979) developed an elimination model in which the choice objects appear as the end nodes of a binary tree, whose interior branches are labeled by aspect weights. That model requires the tree structure to be known in advance, however. The additive tree model of Carroll and DeSoete (1990) allows the tree structure to be estimated from data, which are restricted to pairwise choices.

4. Spatial Models

The tree structures assumed by Tversky and Sattath (1979) are not the only means for coordinating choice objects geometrically. It is often sensible to represent choice alternatives as points x in an n-dimensional space. Pruzansky, Tversky, and Carroll (1982) surmised that perceptual stimuli are adequately represented by multidimensional spatial models, whereas conceptual stimuli are better represented in terms of more discrete structures such as trees.

Böckenholt (1992) and DeSoete and Carroll (1992) have given excellent reviews of probabilistic pair comparison models in which a spatial representation is fundamental. To give a flavor of such models we sketch the wandering vector model presented by DeSoete and Carroll (1986), which is based on earlier ideas suggested by Tucker (1960) and Slater (1960).
The wandering vector model represents choice objects in an \( n \)-dimensional Euclidean space, together with a random vector \( \mathbf{V} \), which fluctuates from trial to trial in a pair comparison experiment and which constitutes an "ideal" direction in the sample space. The vector \( \mathbf{V} \) is assumed to be distributed normally with mean vector \( \mathbf{\mu} \) and covariance matrix \( \Sigma \). A comparison of options \( i \) and \( j \) is determined by three vectors: \( \mathbf{x}_i \) and \( \mathbf{x}_j \), the vector representatives of options \( i \) and \( j \) and \( \mathbf{v} \), a realization of the "wandering" vector \( \mathbf{V} \). Option \( i \) is preferred to option \( j \) whenever the "similarity" of \( i \) to \( j \) as measured by the orthogonal projection of \( \mathbf{x}_i \) on \( \mathbf{v} \) exceeds the corresponding projection of \( \mathbf{x}_j \) on \( \mathbf{v} \). The binary choice probabilities are thus given by

\[
P_{ij} = \text{Prob}(\mathbf{x}_i \cdot \mathbf{V} > \mathbf{x}_j \cdot \mathbf{V}),
\]

where for any vectors \( \mathbf{x} = (x_1, x_2, \ldots x_n) \), \( \mathbf{y} = (y_1, y_2, \ldots y_n) \), the inner product \( \mathbf{x} \cdot \mathbf{y} \) is the number \( x_1y_1 + x_2y_2 + \ldots + x_ny_n \). Using standard theory of the multivariate normal distribution (Ashby, 1992b), one obtains

\[
P_{ij} = \Phi((\mathbf{x}_i - \mathbf{x}_j) \cdot \mathbf{\mu} / \delta_{ij})
\]

Here \( u_i = \mathbf{x}_i \cdot \mathbf{\mu} \) is a (constant) utility associated with option \( i \), and \( \delta_{ij} = (\mathbf{x}_i - \mathbf{x}_j) \cdot \mathbf{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \). Note that the form of Eq. (17) is identical to that of Thurstone's law of comparative judgment (see section II.A.3).

The quantity \( \delta_{ij} \) is a metric that can be interpreted to measure, at least partially, the dissimilarity of options \( i \) and \( j \); Sjöberg (1980) has given some empirical support for this interpretation. On the other hand, there is a considerable body of evidence that empirical judgments of dissimilarity violate the properties required of a metric (Krumhansl, 1978; Tversky, 1977).

A multidimensional similarity model, which appears to address the various shortcomings of the choice models sketched here, is based on the general recognition theory presented by Ashby, Townsend, and Perrin (Ashby & Perrin, 1988; Ashby & Townsend, 1986; Perrin, 1986, 1992). We encounter that theory next in the context of identification.

### B. Identification

#### 1. Ordered Attributes

For stimuli ordered on a one-dimensional continuum, an observer can distinguish perfectly only about seven alternatives spaced equally across the full dynamic range (Miller, 1956). This fact, which is quite robust over different continua, is in sharp contrast to the results of local discrimination experiments of the sort discussed earlier in section II. For example, jnds measured in loudness discrimination experiments employing pure tones
vary from a few decibels at low intensities to a fraction of a decibel at high intensities suggesting, quite contrary to the evidence, that an observer should be able to identify 40 or more tones of increasing loudness spaced evenly over an 80 dB range.

Such puzzling phenomena have prompted a number of authors to study the identification of one-dimensional stimuli as a function of stimulus range (Berliner, Durlach, & Braida, 1977; Braida & Durlach, 1972; Durlach & Braida, 1969; Luce, Green, & Weber, 1976; Luce, Nosofsky, Green, & Smith, 1982; Weber, Green, & Luce, 1977). The data from these studies are accompanied by pronounced sequential effects, first noted by Holland and Lockhead (1968) and Ward and Lockhead (1970, 1971), implicating shifts in response criteria over successive trials and, to a lesser extent, shifts in sensitivity as well (Lacoutre, & Marley, 1995; Luce & Nosofsky, 1984; Marley & Cooke, 1986; Nosofsky, 1983; Treisman, 1985; Treisman & Williams, 1984). Despite the difficulties of interpretation posed by these sequential effects, a robust feature of identification data is the presence of a prominent "edge" effect: Stimuli at the edges of an experimental range are much better identified than stimuli in the middle. As the stimulus range is allowed to increase so that successive stimuli grow farther apart, performance improves, but the edge effect remains. This finding, among others, illustrates that the hope of tying together the data from local psychophysics with those of more global tasks remains an unsettled matter.

2. Multidimensional Stimuli

The basic identification task generates data in the form of a confusion matrix, whose typical entry is the probability $P_{ij}$ of responding stimulus $j$ to the actual presentation of stimulus $i$. A model that has enjoyed considerable success in accounting for such data is the biased choice model (Luce, 1963):

$$P_{ij} = \frac{\beta_i \eta_{ij}}{\sum_j \beta_j \eta_{ij}}.$$

(18)

Here $\eta_{ij}$ is a measure of the similarity of stimuli $i$ and $j$, whereas $\beta_j$ represents a bias toward responding stimulus $j$. Shepard (1957, 1987) has argued that $\eta_{ij} = \exp(-d_{ij})$, where $d_{ij}$ is the distance between alternatives $i$ and $j$ regarded as points in a multidimensional vector space. We have already mentioned that some literature speaks against similarity judgments being constrained by the axioms of a metric (Keren & Baggen, 1981; Krumhansl, 1978; Tversky, 1977). Ashby and Perrin (1988), who favor the general recognition theory (which attempts to account for identification and similarity data within a common multidimensional statistical decision framework), provided additional evidence against the biased choice model.
General recognition theory (GRT) identifies each alternative in an identification experiment with a random vector that takes values in a fixed multidimensional vector space. This vector space is partitioned into disjoint regions, each of which is characteristic of a single response. For illustration, consider the simplest case involving a pair of two-dimensional stimuli, say, A and B, with densities $f_A(x)$ and $f_B(x)$ governing their respective perceptual effects. Statistical decision theory suggests partitioning the two-dimensional sample space on the basis of the likelihood ratio $f_A/f_B$. When the perceptual effects of A and B are jointly normal, curves of constant likelihood ratio are quadratic functions that simplify to lines when the covariance structure of A is the same as that of B (i.e., $\Sigma_A = \Sigma_B$).

Figure 8 generalizes to any number of stimuli varying on any number of dimensions. Figure 9 depicts hypothetical response boundaries for four two-dimensional stimuli labeled by their components: (A1, B1), (A1, B2), (A2, B1), (A2, B2). The response boundaries are chosen so as to maximize accuracy.

General recognition theory yields a conceptually simple expression (though one that is often analytically untractable) for the confusions $P_{ij}$. If $R_j$
is the region of the sample space associated with response stimulus $j$ and $f_i(x)$ is the density governing the perceptual effect of stimulus $i$, then

$$P_{ji} = \int_{B_j} f_i(x) \, dx.$$  

(19)

Numerical methods are normally needed to evaluate such expressions (Ashby, 1992b), in which $f_i(x)$ is multivariate normal.

A competitor to GRT is Nosofsky's generalized context model (GCM), which is an outgrowth of an earlier model of classification proposed by Medin and Shaffer (1978). Unlike GRT, which has its roots in multidimensional statistical decision theory, GCM is based on the idea that people store exemplars in memory as points in a multidimensional space and classify stimuli by proximity in that space to the various exemplars. Nosofsky (1984, 1986) elaborates the model and its assumptions, and in a sequence of articles extends it to take into account phenomena bearing on selective attention (Nosofsky, 1987, 1989, 1991).

These two models, GRT and GCM, seem to account about equally well for a large class of identification and classification data. Because of the different ways each model interprets the same data, a certain amount of scientific controversy has arisen over these interpretations. However, despite their differences in detail, the two models retain much in common,
and one hopes that this fact will promote a third class of models that retains the best features of both GRT and GCM, putting an end to the current disputes.

It has long been thought useful to maintain a distinction between "integral" stimuli—stimuli that are processed as whole entities—and "separable" stimuli—stimuli that are processed in terms of two or more dimensions (see, e.g., Garner, 1974; Lockhead, 1966). Taking such distinctions into account within the framework just presented provides additional and testable constraints on identification data. For a detailed discussion of this and related matters, see Ashby and Townsend (1986), Maddox (1992), and Kadlec and Townsend (1992).

IV. ADDITIVE MEASUREMENT FOR AN ORDERED ATTRIBUTE

In this and the following sections, we shift our focus from models designed to describe the variability of psychophysical data to models that explore more deeply the impact of stimulus structure on behavior. To do so, we idealize response behavior, treating it as if responses exhibit no variability. With few exceptions (e.g., section IV.D), current models do not attempt to combine significant features of both stimulus structure and variable response behavior. It has proved very difficult to combine both phenomena in a single approach due to, in our opinion, the lack of a qualitative theory of randomness.

A. Ranking and Order

Stimuli can be ordered in a variety of ways ranging from standard physical procedures—ordering masses by, say, an equal-arm pan balance or tones by physical intensity (e.g., decibels)—to subjective attributes—perceived weight, perceived loudness, preference among foods, and so on. In each case, the information that is presumed to exist or to be obtainable with some effort is the order between any two objects in the domain that is established by the attribute. Let $A$ denote the domain of stimuli and let $a$ and $b$ be two elements of $A$, often written $a, b \in A$. Then we write $a \succeq b$ whenever $a$ exhibits at least as much of the attribute as does $b$.

The order $\succeq$ can be established either by presenting pairs and asking a subject to order them by having the subject rank order the entire set of stimuli, by rating them in some fashion, or by indirect methods some of which we describe shortly.

Of course, as we observed in section II, for most psychological attributes such consistency is, at best, an idealization. If you ask a subject to order $a$ and $b$ more than once, the answer typically changes. Indeed, one assumes
that, in general, a probability $P_{a,b}$ describes the propensity of a subject to order $a$ and $b$ as $a \succeq b$. There are ways to induce an order from such probabilities. One is simply to use the estimated propensity as the source of ordering, namely,

$$a \succeq b \text{ holds if and only if } P_{a,b} \succeq 1.$$  \hfill (20)

If a Fechnerian model holds (Section II.A.1), this is the order established by the underlying subjective scale $u$. Another order, which often is of considerable importance, is not on the stimuli themselves, but on pairs of them:

$$(a, b) \succeq (a', b') \text{ whenever } P_{a,b} \succeq P_{a',b'}.$$  \hfill (21)

Still another way to establish a psychological order is by measuring the time it takes a subject to decide whether $a$ has more of the attribute than $b$. If $L_{a,b}$ denotes the mean response time for that judgment, then replacing $P$ in Eq. (21) with $L$ yields a potentially new order that is well defined. In practice, these two orders are not wholly independent; witness the existence of speed-accuracy trade-offs. Some authors have conjectured that $L_{a,b}$ may be a decreasing function of $|P_{a,b} - 1|$; however, nothing really simple seems to hold (Luce, 1986).

The purpose of this section and the subsequent two sections is to study some of the properties of orders on structures and certain numerical representations that can arise. This large and complex topic has been treated in considerable detail in several technical sources: Falmagne (1985); Krantz, Luce, Suppes, and Tversky (1971); Luce, Krantz, Suppes, and Tversky (1990); Narens (1985); Pfanzagl (1971); Roberts (1979); and Suppes, Krantz, Luce, and Tversky (1989); and Wakker (1989). For philosophically different approaches and commentary, see Deacon, Onghena, and Janssen (1995), Ellis (1966), Michell (1990, 1995), Niederée (1992, 1994), and Savage and Ehrlich (1992).

1. Transitivity and Connectedness

To the extent that an order reflects an attribute that can be said to exhibit "degree of" or "amount of," then we expect it to exhibit the following property known as transitivity: For all $a, b, c \in A$,

$$\text{if } a \succeq b \text{ and } b \succeq c, \text{ then } a \succeq c.$$  \hfill (22)

Transitivity is a property of numbers: $12 \succeq 8$ and $8 \succeq 5$ certainly means that $12 \succeq 5$. At various times we will focus on whether transitivity holds.

A second observation is that for very many attributes it is reasonable to assume the following property, which is known as connectedness: For all $a$ and $b \in A$,

$$\text{either } a \succeq b \text{ or } b \succeq a \text{ or both.}$$  \hfill (23)

The orders defined by Eqs. (20) and (21) obviously satisfy connectedness.
When both \( a \succeq b \) and \( b \succeq a \) hold, we write \( a \sim b \) meaning that \( a \) and \( b \) are indifferent with respect to the attribute of the ordering. It does not usually correspond to equality; two objects can have the same weight without being identical. If \( a \succeq b \) but not \( a \sim b \), then we write \( a > b \). Whatever the attribute corresponding to \( \succeq \) is called, the attribute corresponding to \( > \) receives the same name modified by the adjective \emph{strict}. Equally, if \( > \) has a name, \( \preceq \) is prefixed by \emph{weak}. So, for example, if \( > \) denotes preference, \( \preceq \) denotes weak preference.

Should one confront an attribute for which connectedness fails, so that for some \( a \) and \( b \) neither \( a \succeq b \) nor \( b \succeq a \), we usually speak of \( a \) and \( b \) as being \emph{noncomparable} in the attribute and the order as being \emph{partial}. For example, suppose one were ordering a population of people by the attribute “ancestor of.” This is obviously transitive and equally obviously not connected. All of the attributes discussed in this chapter are assumed to be connected.

A connected and transitive order is called a \emph{weak order}. When indifference, \( \sim \), of a weak order is actually equality, that is, \( a \sim b \) is equivalent to \( a = b \), the order is called \emph{simple} or \emph{total}. The numerical relation \( \succeq \) is the most common example of a simple order, but very few orders of scientific interest are stronger than weak orders unless one treats classes of equivalent elements as single entities.

### 2. Ordinal Representations

One major feature of measurement in the physical sciences and, to a lesser degree, in the behavioral and social sciences is the convenience of representing the order information numerically. In particular, it is useful to know when an empirical order has an \emph{order preserving} numerical representation, that is, when a function \( \phi \) from \( A \) into the real numbers \( \mathbb{R} \) (or the positive real numbers \( \mathbb{R}_+ \)) exists such that for all \( a, b \in A \),

\[
    a \succeq b \text{ is equivalent to } \phi(a) \succeq \phi(b).
\]

Because \( \succeq \) is a total order, it is not difficult to see that a necessary condition is that \( \succeq \) be a weak order. When \( A \) is finite, being a weak order is also sufficient because one can simply take \( \phi \) to be the numerical ranking: assign 1 to the least element, 2 to the next, and so on. For infinite structures, another necessary condition must be added to achieve sufficiency; it says, in effect, that \((A, \succeq)\) must contain a subset that is analogous to the rational numbers in the reals, that is, a countable order-dense subset. The details, listed as the Cantor or Cantor-Birkhoff theorem, can be found in any book on the theory of measurement (e.g., Krantz et al., 1971, section 2.1, or Narens, 1985, p. 36).

One feature of Eq. (24) is that if \( f \) is any strictly increasing function\(^6\) from \( \mathbb{R} \) to \( \mathbb{R} \),\(^6\) then \( f(\phi) \) is an equally good representation of \((A, \succeq)\). When a

\[^6\] If \( x > y \), then \( f(x) > f(y) \).
representation has this degree of nonuniqueness, it is said to be of ordinal scale type. One drawback of this nonuniqueness is that little of arithmetic or calculus can be used in a way that remains invariant under admissible scale changes. For example, if \( \phi \) is defined on the positive real numbers, then \( f(\phi) = \phi^2 \) is an admissible transformation. If \( \phi(a) = 5, \phi(b) = 4, \phi(c) = 6, \) and \( \phi(d) = 3, \) then \( \phi(a) + \phi(b) \equiv \phi(c) + \phi(d), \) but \( \phi^2 \) reverses the order of the inequality. Therefore great care must be taken in combining and compressing information that is represented ordinally (see section VI).

3. Nontransitivity of Indifference and Weber’s Law

A very simple consequence of \( \succeq \) being a weak order is that both the strict part, \( >, \) and the indifference part, \( \sim, \) must also be transitive. Although the transitivity of \( > \) seems plausible for many attributes, such may not be the case for \( \sim, \) if for no other reason than our inability to discriminate very small differences. The measurement literature includes a fair amount of material on orderings for which \( > \) is transitive and \( \sim \) is not. Conditions relating them are known that lead to a representation in terms of two numerical functions \( \phi \) and \( \delta, \) where \( \delta > 0 \) is thought of as a threshold function:

\[
a > b \text{ is equivalent to } \phi(a) \geq \phi(b) + \delta(b), \quad (25a)
\]

\[
a \sim b \text{ is equivalent to } \phi(a) - \delta(a) < \phi(b) < \phi(a) + \delta(a). \quad (25b)
\]

Orders exhibiting such a threshold representation are known as semiorders and interval orders, the latter entailing different upper and lower threshold functions (Fishburn, 1985; Suppes et al., 1989, chap. 16).

One major question that has been studied is, when is it possible to choose \( \phi \) in such a way that \( \delta \) is a constant? This question is very closely related to the psychophysical question of when does Weber’s law (just detectable differences proportional to stimulus intensity) hold in discrimination.

To be specific, in a context of probabilistic responses, suppose a probability criterion \( \lambda, \frac{1}{2} \leq \lambda < 1 \) (e.g., 0.75 is a common choice) is selected to partition the discriminable from the indiscriminable. This defines the algebraic relation \( \succeq_\lambda \) in terms of the probabilities by

\[
a >_\lambda b \text{ is equivalent to } P_{a,b} \geq \lambda \quad (26a)
\]

\[
a \sim_\lambda b \text{ is equivalent to } 1 - \lambda < P_{a,b} < \lambda. \quad (26b)
\]

It can be shown that for \( \succeq_\lambda \) to have a threshold representation, Eq. (25), with a constant threshold, is equivalent to the Weber function of \( P \) satisfying Weber’s law (see sections II.A.2 and VII.B.1).
B. Conjoint Structures with Additive Representations

Let us return to the simpler case of a weak order. The ordinal representation is rather unsatisfactory because of its high degree of nonuniqueness, so one is led to consider situations exhibiting further empirical information that is to be represented numerically. That consideration is the general topic of this subsection and section IV.C.

1. Conjoint Structures

The sciences, in particular psychology, are replete with attributes that are affected by several independent variables. For example, an animal's food preference is affected by the size and composition of the food pellet as well as by the animal's delay in receiving it; the aversiveness of an electric shock is affected by voltage, amperage, and duration; loudness depends on both physical intensity and frequency; the mass of an object is affected by both its volume and the density of material from which it is made; and so forth.

Each of these examples illustrates the fact that we can and do study how independently manipulable variables trade off against one another in influencing the dependent attribute. Thus, it is always possible to plot those combinations of the variables that yield equal levels of the attribute. Economists call these indifference curves, and psychologists have a myriad of terms depending on context: ROCs for discrimination (section II.B.1), equal-loudness contours, and curves of equal aversiveness, among others.

The question is whether these trade-offs can be a source of measurement. Let us treat the simplest case of two independent variables; call their domains $A$ and $U$. Thus a typical stimulus is a pair, denoted $(a, u)$, consisting of an $A$ element and a $U$ element. The set of all such pairs is denoted $A \times U$. The attribute in question, $\succ$, is an ordering of $A \times U$, and we suppose it is a weak order. One possibility for the numerical representation $\phi$ is that in addition to being order preserving, Eq. (24), it is additive over the factors $A$ and $U$, meaning that there is a numerical function $\phi_A$ on $A$ and another one $\phi_U$ on $X$ such that

$$\phi(a, u) = \phi_A(a) + \phi_U(u),$$

that is, for $a, b \in A$ and $u, v \in U$,

$$(a, u) \preceq (b, v) \text{ is equivalent to } \phi_A(a) + \phi_U(u) \preceq \phi_A(b) + \phi_U(v).$$

Witness the shape of equal loudness contours at low intensity, which is the reason for loudness compensation as well as intensity controls on audio amplifiers.

The terms independent variable, factor, and component are used interchangeably in this literature, except when component refers to the level of a factor.

---

7 Witness the shape of equal loudness contours at low intensity, which is the reason for loudness compensation as well as intensity controls on audio amplifiers.

8 The terms independent variable, factor, and component are used interchangeably in this literature, except when component refers to the level of a factor.
Such additive representations are typically used in the behavioral and social sciences, whereas the physical sciences usually employ a multiplicative representation into the positive real numbers, $\mathbb{R}_+$. The multiplicative representation is obtained from Eq. (27) by applying an exponential transformation, thereby converting addition to multiplication: $e^{x+y} = e^x e^y$.

2. The Existence of an Additive Representation

Two questions arise: What must be true about $\succeq$ so that an additive representation, Eq. (27), exists; and if one does exist, how nonunique is it? Mathematically precise answers to these questions are known as representation and uniqueness theorems. It is clear that for such a strong representation to exist, the qualitative ordering $\succeq$ must be severely constrained. Some constraints are easily derived. For example, if we set $u = v$ in Eq. (27b), we see that because $\Phi_i(u)$ appears on both sides of the inequality it can be replaced by any other common value, for example, by $\Phi_j(w)$. Thus, a necessary qualitative condition, known as independence in this literature, is that for all $a, b$ in $A$ and $u, v$ in $U$,

\[(a, u) \succeq (b, u) \text{ is equivalent to } (a, v) \succeq (b, v). \tag{28a}\]

Similarly, one can hold the first component fixed and let the second one vary:

\[(a, u) \succeq (a, v) \text{ is equivalent to } (b, u) \succeq (b, v). \tag{28b}\]

For a long time psychologists have been sensitive to the fact that Eq. (28) necessarily holds if the attribute has an additive representation, and in plots of indifference curves (often with just two values of the factors) the concern is whether or not the curves "cross." Crossing rejects the possibility of an additive representation; however, as examples will show, the mere fact of not crossing is insufficient to conclude that an additive representation exists. The reason is that other conditions are necessary beyond those that can be deduced from weak ordering and Eq. (28).

For example, suppose we have two inequalities holding with the property that they have a common $A$ value in the right side of the first qualitative inequality and the left side of the second one and also a common $U$ value in the left side of the first and the right side of the second qualitative inequalities; then when the corresponding numerical inequalities are added and the common values canceled from the two sides one concludes from Eq. (27) that

\[\text{if } (a, u) \succeq (g, u) \text{ and } (g, u) \succeq (b, u) \text{ then } (a, u) \succeq (b, u). \tag{29}\]

9 The term uniqueness is used despite the fact that the thrust of the theorem is to tell just how nonunique the representation is.

10 Mathematically, it might better be called monotonicity, but the term independence is widely used.
1 The Representational Measurement Approach to Problems

This property is known as *double cancellation* because, in effect, two values, \( w \) and \( g \), are canceled. In recent years it has been recognized that at least both Eqs. (28) and (29) need to be checked when deciding if an additive representation is possible (see Michell, 1990).

It is not difficult to see that we can go to three antecedent inequalities with an appropriate pattern of common elements so that Eq. (27) leads to further, more complex conditions than Eq. (29). Do we need them all?

The answer (see Chap. 9 of Krantz et al., 1971) is yes if we are dealing with a finite domain \( A \times U \). For infinite domains, however, it turns out that the properties of weak order, independence, and double cancellation—that is, Eqs. (22), (23), (28), and (29)—are sufficient for those (important) classes of structures for which the independent variables can reasonably be modeled as continuous (often physical) variables. Such continuous models are typically assumed in both psychophysics and utility theory. The added conditions are a form of solvability of indifferences (see the next subsection) and a so-called *Archimedean property*, which we do not attempt to describe exactly here (see section IV.C and chap. 6 of Krantz et al., 1971). Suffice it to say that it amounts to postulating that no nonzero interval is infinitesimal relative to any other interval; all measurements are comparable.

3. The Uniqueness of Additive Representations

The second question is, how nonunique is the \( \phi \) of Eq. (27)? It is easily verified that \( \Psi_A = r \Phi_A + s \) and \( \Psi_U = r \Phi_U + t \), where \( r > 0 \), \( s \), and \( t \) are real constants, is another representation. Moreover, these are the only transformations that work. Representations unique up to such positive affine (linear) transformations are said to be of *interval scale type* (Stevens, 1946, 1951).

4. Psychological Applications

Levelt, Riemersma, and Bunt (1972) collected loudness judgments over the two ears and constructed an additive conjoint representation. Later Gigrenzer and Strube (1983), using an analysis outlined by Falmagne (1976) that is described in section IV.D, concluded that additivity of loudness fails, at least when one of the two monaural sounds is sufficiently louder than the other: the louder one dominates the judgments. To the extent that additivity fails, we need to understand nonadditive structures (section VI).

Numerous other examples can be found in both the psychological and marketing literatures. Michell (1990) gives examples with careful explanations.

11 Of course, any experiment is necessarily finite. So one can never test all possible conditions, and it is a significant inductive leap from the confirmation of these equations in a finite data set to the assertion that the properties hold throughout the infinite domain. For finite domains one can, in principle, verify all of the possible cancellation properties.
FIGURE 10 The construction used to create an operator \( Q \) on one component of a conjoint structure that captures the trade-off of information. Pairs connected by dashed lines that intersect in the middle are equivalent. Panel (a) schematically represents the components as continua with distinguished points \( a_0 \) and \( u_0 \). Panel (b) maps the interval \( a,b \) to the interval \( u_0,\pi(b) \). Panel (c) illustrates adding the latter interval to \( a_0 \) to get the "sum" interval \( a_0,\pi(b) \).

C. Concatenation Structures with Additive Representations

1. Reducing the Conjoint Case to a Concatenation Operation

It turns out that the best way to study independent conjoint structures, whether additive or not (section VI.B.2), is to map all of the information contained in \( (A \times U, \succeq) \) into an operation and ordering on \( A \). Consider the following definition:

\[
\begin{align*}
a \succeq_A b \text{ is equivalent to } (a, u) \succeq (b, u) & \quad (30)
\end{align*}
\]

Independence, Eq. \( (28) \), says that the order induced on \( A \) by Eq. \( (30) \) is unaffected by the choice of \( u \) on the second factor.

Inducing an operation on \( A \) is somewhat more complex. The general procedure is outlined in Figure 10 and details are given in Krantz et al. (1971, p. 258). It rests first on arbitrarily picking an element from each factor, say, \( a_0 \) from \( A \) and \( u_0 \) from \( U \). Next, one maps what intuitively can be thought of as the "interval" \( a,b \) of the \( A \) component onto an equivalent "interval" of the \( U \) component, which we will call \( u_0,\pi(b) \). The formal definition is that \( \pi(b) \) satisfies the equivalence:
Clearly, one must make an explicit assumption that such a solution $\pi(b)$ can always be found. Such a solvability condition is somewhat plausible for continuous dimensions and far less so for discrete ones.

The third step is to "add" the interval $a,b$ to the interval $a,a$ by first mapping the interval $a,b$ to $w_0\pi(b)$, Eq. (31), and then mapping that interval back onto the interval from $a$ to a value called $aO_A b$ that is defined as the solution to the indifference

$$ (aO_A b, w_0) \sim (a, \pi(b)). $$

The operation $O_A$ is referred to as one of concatenation or "putting together." It turns out that studying the concatenation structure $(A, \geq_A, O_A, a_0)$ is equivalent to studying $(A \times X, \geq)$, and because of its importance in physical measurement it is a well-studied mathematical object (see Krantz et al., 1971, chap. 3; Narens, 1985, chap. 2). For simplicity, let us drop the $A$ subscripts and just write $(A, \geq, O, a_0)$.

2. Properties of $O$ and $\geq$

It is clear that those concatenation structures arising from additive conjoint ones will involve some constraints on $O$ and on how it and $\geq$ are related. It is not terribly difficult to show that independence of the conjoint structure forces the following monotonicity property:

$$ a \geq a' \text{ is equivalent to } a \circ b \geq a' \circ b \text{ and } $$

$$ b \geq b' \text{ is equivalent to } a \circ b \geq a \circ b'. $$

Intuitively, these conditions are highly plausible: Increasing either factor of the operation increases the value. The double-cancellation property implies the following property of $O$, which is called associativity:

$$ a \circ (b \circ c) \sim (a \circ b) \circ c. $$

A third property is that $a_0$ acts like a "zero" element:

$$ a_0 \circ a \sim a \circ a_0 \sim a. $$

Solvability ensures for every $a, b \in A$ not only that $a \circ b$ is defined, but that each element $a$ has an inverse element $a^{-1}$ with the property:

$$ a \circ a^{-1} \sim a^{1-1} \circ a \sim a_0. $$

Finally, we formulate the Archimedean property of such a structure. By repeated applications of Eq. (34), it does not matter which sequence of binary groupings is used in concatenating $n$ elements. Let $a(n)$ denote $n$ concatenations of $a$ with itself. Suppose $a > a_0$. The Archimedean assumption says that for any $b \in A$ one can always find $n$ sufficiently large so that $a(n) > b$. Intuitively, this means that $a$ and $b$ are commensurable.
3. H{"o}lder’s Theorem

In 1901 the German mathematician O. H{"o}lder proved a version of the following very general result (H{"o}lder, 1901/1996). Suppose a structure \((A, \preceq, \circ, a_0)\) satisfies the following properties: \(\preceq\) is a weak order, monotonicity, associativity, identity, inverses, and Archimedeaness, Eqs. (22), (23), and (33) through (36). Such structures are called Archimedean ordered groups, the term group encompassing the three properties of Eqs. (34) through (36) (see footnote 2). The result is that such a structure has a representation \(\phi\) into the additive real numbers, which means that \(\phi\) is order preserving, Eq. (24), and additive over \(\circ\):

\[
\phi(a \circ b) = \phi(a) + \phi(b). \tag{37}
\]

From this fact, it is fairly easy to establish that a conjoint structure satisfying independence, Eq. (28), double cancellation, Eq. (29), solvability, and a suitable Archimedean condition has an additive representation.

4. Ratio Scale Uniqueness

The nonuniqueness of H{"o}lder’s additive representation is even more restricted than that for conjoint structures: \(\phi\) and \(\psi\) are two additive representations if and only if for some constant \(r > 0\), \(\psi = r\phi\). The class of such representations is said to form a ratio scale (Stevens, 1946, 1951). A conjoint representation is interval rather than ratio because the choice of the “zero” element \(a_0\) is completely arbitrary; however, in H{"o}lder’s additive structure the concatenation of the zero element with another element leaves the latter unchanged and so \(\phi(a_0) = 0\) in all representations \(\phi\).

5. Counting Yields the Representation

The key to the construction of a representation is to fix some element \(c > a_0\) and find the number of copies of \(c\) that are required to approximate any element \(a\). This is done by considering \(m\) copies of \(a\) and using the Archimedean axiom to find the smallest \(n\), which is a function of \(m\), such that \(c(n - 1) + a > c(n)
\]

One shows that \(n/m\) approaches a limit as \(m\) approaches \(\infty\) and defines \(\phi(a)\) be that limit where, of course, \(\phi(c) = 1\). Next one proves that such limits are additive over \(\circ\).

6. Extensive Structures

If one looks just at the positive part of \(A\), that is, \(A_+ = \{a: a \in A \text{ and } a > a_0\}\) with \(\preceq\) and \(\circ\) restricted to that part, one has the added feature that for all \(a, b \in A_+\), both \(a \circ b > a\) and \(a \circ b < b\), and \(a > b\) implies \(a > a_0\) such that \(a \circ c > b\) (\(c = a \circ b^{-1}\)). Such a structure is called extensive (in contrast to the intensive structures discussed in section V). One can use H{"o}lder’s theorem
1 The Representational Measurement Approach to Problems

37
to show that any extensive structure has a ratio scale representation into \((\mathbb{R}_+, \geq, +)\), the ordered, additive, positive real numbers.

These structures were, in fact, the first to have been formalized as models of certain basic types of physical measurement. For example, if \(A\) denotes straight rods, with \(\geq\) the qualitative ordering of length obtained by direct comparison and \(\circ\) the operation of abutting rods along a straight line, then Eqs. (22), and (23), and (33) through (36) are all elementary physical laws. Mass, charge, and several other basic physical quantities can be measured in this fashion.

Aside from their indirect use in proving the existence of additive conjoint measurement representations, extensive structures have played only a limited descriptive role in the behavioral and social sciences, although they can serve as null hypotheses that are then disconfirmed. It is not that there is a dearth of operations but rather that one or more of Eqs. (33) through (36) usually fail, most often associativity, Eq. (34). For example, various forms of averaging, although involving + in their representation, are not associative (see section V). Receiving two goods or uncertain alternatives is an operation of some importance in studying decision making, and it is unclear at present whether or not it is associative.

7. Combining Extensive and Conjoint Structures

Often the components of a conjoint structure \((A \times U, \geq)\) are themselves endowed with empirical operations \(*_A\) and/or \(*_U\) that form extensive structures on \(A\) and \(U\), respectively. Many physical examples exist, for example, mass and velocity ordered by kinetic energy, as well as psychological ones such as sound intensities to the two ears. An important question is, how are the three structures interrelated?

One relation of great physical importance is the following distribution law: For all \(a, b, c, d \in A, u, v \in U,\)

if \((a, u) \sim (c, v)\) and \((b, u) \sim (d, v),\) then \((a*_{A}b, u) \sim (c*_{A}d, v),\) (38)

and a similar condition is true for the second component. It has been shown that if \(\phi_A\) and \(\phi_U\) are additive representations of the two extensive structures, then there is a constant \(\beta\) such that

\[
\phi_A \phi_U^\beta
\]

(39)
is a multiplicative representation of \((A \times U, \geq)\). (Luce et al., 1990, summarize the results and provide references to the original literature.) The exponent \(\beta\) characterizes the trade-off between the two extensive measures. For example, in the case of kinetic energy \(\beta\) is 2, which simply says a change in velocity by a factor \(k\) is equivalent to a change in mass by a factor \(k^2\). Such trade-off connections as given by Eq. (39) are common in physical measurement, and their existence underlies the dimensional structure of classical
physical measurement. Moreover, their existence, as embodied in Eqs. (38) and (39), is also the reason that physical units are always products of powers of several basic extensive measures, for example, the unit of energy, the erg, is g cm²/s². (For details see chap. 10 of Krantz et al., 1971, and chap. 22 of Luce et al., 1990.)

Luce (1977) also studied another possible relation between an additive conjoint structure whose components are also extensive. Let \( a(j) \) denote \( j \) concatenations of \( a \), and suppose there exist positive integers \( m \) and \( n \) such that for all positive integers \( i \) and \( a, b \in A \) and \( u, v \in U \),

\[
(a, u) \sim (b, v) \Rightarrow (a(i^m), u(i^n)) \sim (b(i^m), v(i^n)). \tag{40}
\]

Under some assumptions about the smoothness of the representations, it can be shown that the representation is of the form:

\[
r_A \phi_A^{a(i^m)} + r_U \phi_U^{u(i^n)} + s,
\]

where \( r_A > 0 \), \( r_U > 0 \), and \( s \) are real constants. For example, the Levelt et al. (1972) conjoint analysis of loudness judgments over the two ears supported not only additivity but the power functions of Eq. (41); however, see section IV.D. The power functions arising in both Eqs. (39) and (41) are psychologically interesting because, as is discussed in section VII.C, substantial empirical evidence exists for believing that many psychological attributes are approximately power functions of the corresponding physical measures of intensity. Of course, as noted earlier, later studies have cast doubt on the additivity of loudness between the ears.

D. Probabilistic Conjoint Measurement

The variability that accompanies psychophysical data rules out the possibility of direct empirical tests of algebraic measurement axioms. Probabilistic versions of both extensive measurement (Falmagne, 1980) and conjoint measurement (Falmagne, 1976) have been proposed, although as we shall see they often exhibit the difficulty alluded to in section I. We treat only the conjoint case here.

Consider the discrimination of pure tones \((a, u)\) presented binurally: \( a \) denotes the intensity of a pure tone presented to the left ear of an observer, and \( u \) is the intensity of the same frequency presented, in phase, to the right ear. The data are summarized in terms of the probability \( P_{a, b} \) that the binaural stimulus \( (a, u) \) is judged at least as loud as stimulus \((b, v)\).

One class of general theories for such data that reflects the idea that the stimuli can be represented additively asserts that

\[
P_{a, b} = H[l(a) + r(u), l(b) + r(v)] \tag{42}
\]

for some suitable functions \( l, r \), and \( H \) with \( H(u, u) = \frac{1}{2} \).
To generate a loudness match between two binaural tones, one fixes three of the monaural components, say, \(a\), \(u\), and \(v\), and seeks \(b\) such that \(P_{au,bv} = \frac{b}{l}\). Of course, this must be an estimate and is therefore subject to variability. This suggests replacing the deterministic, but empirically unattainable prediction from Eq. (42) that
\[
(l(a) + r(u) = l(b) + r(v)
\]
by a more realistic one that substitutes a random variable \(U_{au}\) for \(b:\)
\[
l(U_{au}) = l(a) + r(u) - r(v) + E_{au},
\]
where \(E_{au}\) is a random error.

This proposal illustrates the difficulty in simultaneously modeling structure and randomness. The assumption that the error is additive in the additively transformed data seems arbitrary and is, perhaps, unrealistic. Certainly, no justification has been provided. One would like to see Eq. (44) as the conclusion of a theorem, not as a postulate. Of course, writing equations like Eq. (44) is a widespread, if dubious, tradition in statistics.

If the random error \(E_{au}\) is assumed to have median zero, then the random representation Eq. (44) simplifies to the deterministic Eq. (43) upon taking medians over the population. This suggests studying the properties of the function \(m_{au}(a) = \text{Median}(U_{au})\). Falmagne (1976) showed, in the context of natural side conditions, that if the medians satisfy the following property of cancellation,
\[
l(a) + r(u) + r(v) = l(b) + r(u) + r(v)
\]
then they can be represented in the additive form of Eq. (43): \(l(a) + r(u) = l(m_{au}(a)) + r(v)\). The linkage between the median functions and algebraic conjoint measurement is provided by a relation \(\leq\) over the factor pairs:
\[
uw \leq bv\text{ is equivalent to } m_{uw}(a) \leq b.
\]
It is straightforward to show that \(\leq\) is a weak order and that cancellation, Eq. (45), implies the double cancellation condition, Eq. (29), of conjoint measurement.

Falmagne found the cancellation condition to be supported by empirical data, in agreement with earlier work of Levelt, Riemersma, and Bunt (1972). Later, however, Gigerenzer and Strube (1983) showed that cancellation breaks down when one of the monaural components sufficiently dominates the other.

Special cases of Eq. (42), namely,
\[
P_{au,bv} = F[l(a) + r(u) - l(b) - r(v)],
\]
\[
P_{au,bv} = F\left(\frac{l(a) + r(u) - l(b) + r(v)}{l(b) + r(v)}\right),
\]

are considered in the next section.
were investigated by Falmagne, Iverson, and Marcovici (1979) and Falmagne and Iverson (1979). Of these, the second, Eq. (47b), was found to provide the best account of discrimination data. Note that this additive-ratio model suggests an alternative form to Eq. (44), namely,

\[ l(a) + r(a) = \left[ l(U_{aw}) + r(v) \right] E_{aw}. \] (48)

Once again, it should be apparent that the theoretical representation of both structure and randomness seems to rest on arbitrary assumptions before the analysis can proceed.

E. Questions

The reader probably has some unresolved questions about the measurement structures we have examined. One likely question centers on the curious fact that these examples have included only three scale types: ordinal, interval, and ratio. The psychologist S. S. Stevens (1946, 1951) was the first to point out that only these scale types seemed to play a serious role in the physical sciences. Is there a reason for this? During the 1980s it was discovered that, indeed, there is a fairly deep reason, which we examine in section VI.

A second major question for psychologists is whether measurement is limited to additive conjoint and extensive structures. Additivity is clearly too restrictive for the behavioral and social sciences. Do we understand nonadditive structures well enough for them to be useful in science? The answer is yes, and the general theory is outlined partly in section V and more fully in section VI.

V. WEIGHTED-AVERAGE MEASUREMENT FOR AN ORDERED ATTRIBUTE

In addition to the attributes endowed with concatenation operations that have additive representations, other important attributes have averaging representations. To illustrate the scientific range we cite four examples:

- Numerical weighted averages, such as

\[ \mu x + (1 - \mu) y, \] (49)

where the weight \( \mu \in (0, 1) \), abound in statistical data processing.

- When two physical objects at different temperatures are permitted to come into equilibrium with each other, their final temperature is a weighted average of the two initial temperatures with weights that depend on the compositions and volumes of the objects.
Psychophysicists sometimes study subjective attributes, such as brightness or loudness, by having observers produce stimuli that lie a specified fraction along the subjective interval between two experimenter presented stimuli. When the fraction is $\frac{1}{2}$, the method is called bisection. We denote the bisection point of stimuli $a$, $b$ either by function notation $F(a, b)$ or operator notation $a \circ b$.

In theories of decision making, uncertain alternatives (or gambles) are alternatives consisting of outcomes determined by chance events arising in some "experiment." In the binary case, $a \circ b$ denotes the uncertain alternative in which consequence $a$ is received if event $E$ occurs and $b$ otherwise when the "experiment" is conducted. The simplest example is a lottery. An important class of theories for such binary gambles, known generically as subjective expected utility (SEU), takes the form of weighted average representations.

Clearly, understanding those structures leading to averaging representations is of broad scientific interest. This section describes some of what we know about these structure and their applications.

A. Binary Intensive Structures

1. Betweenness, Idempotence, and Bisymmetry

Consider again a structure $A = (A, \succeq, \circ)$, where $\succeq$ is a binary relation and $\circ$ a binary operation on $A$. Our first concern is to develop properties that $A$ must exhibit in order to have a weighted average representation. As before, weak ordering, Eqs. (22) and (23), and monotonicity, Eq. (33), must be satisfied. But it is easy to see that other properties of extensive structures fail.

For example, when $\circ$ has an averaging representation, $\circ$ cannot be positive in the sense that both $a \circ b > a$ and $a \circ b > b$, but rather a betweenness property holds: For all $a, b, \in A$ with $a \succeq b$,

$$a \succeq a \circ b \succeq b \text{ and } a \succeq b \circ a \succeq b,$$

from which it follows immediately that for all $a \in A$,

$$a \circ a \succeq a,$$

which property is called idempotence.

Another property that fails is associativity, Eq. (34); it is replaced by a relationship among four, not three, elements. Using Eq. (49), consider the separate weighted averages of $w$, $x$ and $y$, $z$. Then the weighted average of these two averages is easily seen to be

\[12\] The term experiment is used here in the special sense of statistics, not in the usual psychological sense.
Observe that because \( x + y = y + x \), this is the same as averaging the pairs \( w, y \) and \( x, z \) and then averaging the results. Therefore, the qualitative analog is

\[
(a \circ b) \circ (c \circ d) \sim (a \circ c) \circ (b \circ d),
\]

a property known as bisymmetry.

2. The Representation and Reduction to Conjoint Measurement

A concatenation structure \( \mathcal{S} \) that satisfies weak ordering, monotonicity, and betweenness is called intensive. One can show that an intensive structure that is bisymmetric, suitably solvable, and suitably Archimedean has a numerical representation \( \phi \) and a constant \( \mu \in (0, 1) \) such that for all \( a, b \in A \)

\[
\phi(a \circ b) = \mu \phi(a) + (1 - \mu) \phi(b).
\]

One way to prove this is to define \( \succeq' \) on \( A \times A \) by:

\[
(a, c) \succeq' (b, d) \text{ if and only if } a \circ c \succeq b \circ d.
\]

One then shows that the assumed properties are sufficient to prove that the conjoint structure\(^{13} \) \((A \times A, \succeq)\) satisfies the conditions of section IV.B, so it has an additive representation. Of course, there were two functions in that representation, see Eq. (27), but in this case, where \( U = A \), they can be proved to be proportional.

As with conjoint measurement, the representation of Eq. (53) is of interval scale type.

From an empirical point of view, the upshot of this result is that idempotence and bisymmetry are the key properties distinguishing a structure having an averaging representation from an associative structure with an additive representation.

Probably the most extensive uses of averaging operations in psychology are found in three areas: statistics, individual decision making (section V.C), and applications of a method called functional measurement (section V.D). Although binary models are of some interest in these applications, the most interesting cases involve combining more than two entities at a time. We turn to those cases in the next section before taking up applications.

B. Generalized Averaging

One can continue to think of combining more than two things as an operation, but usually we use functional notation. Therefore the structure is \((A,\)

\(^{13}\) This is the special case of the earlier conjoint structures with \( U = A \).
The Representational Measurement Approach to Problems

\( \geq, F \), where \( F: A \times \ldots \times A \rightarrow A \) (just as \( \bigcirc: A \times A \rightarrow A \); so a typical result from the operation is written as \( a = F(a_1, \ldots, a_n) \). It is not difficult to see that monotonicity generalizes to saying that for any component \( i \),

\[
a'_i \geq a_i \text{ is equivalent to } F(a_1, \ldots, a_i, \ldots, a_n) \geq F(a_1, \ldots, a_i, \ldots, a_m).
\]

Idempotence becomes

\[
F(a, \ldots, a, \ldots) = a.
\]

There is no obvious generalization of bisymmetry, but we do not really need one because it suffices to require that bisymmetry hold on any pair of components for any arbitrary, but fixed, choice for the remaining \( n-2 \).

The resulting representation on assuming weak order, monotonicity, idempotence, and suitable solvability and Archimedean properties is the existence of \( \phi: A \rightarrow \mathbb{R} \) and \( n \) numerical constants \( \mu_i \) such that

\[
\mu_i \geq 0, \quad i = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^{n} \mu_i = 1 \quad (55a)
\]

\[
\phi[F(a_1, \ldots, a_n)] = \sum_{i=1}^{n} \mu_i \phi(a_i). \quad (55b)
\]

C. Utility of Uncertain Alternatives

1. Subjective Expected Utility (SEU)

For decision making, the model is somewhat more complex and developing proofs is considerably more difficult. But the representation arrived at is easy enough to state. A typical uncertain alternative \( g \) assigns consequences to events; that is, it is an assignment \( g(E_i) = g_i, i = 1, \ldots, n \), where the \( E_i \) are disjoint events taken from a family of events and the \( g_i \) are consequences from a set of possible consequences (e.g., amounts of money). Lottery examples abound in which the events are often sets of ordered \( k \)-tuples of numbers, where \( k \) usually is between 3 and 10, to a few of which award amounts are assigned.\(^{14}\) The SEU theories in their simplest form assert that for a sufficiently rich collection of events and uncertain alternatives there is a (subjective) probability measure \( \mu \) over the family of events.,\(^{15}\) and there is

\(^{14}\) For example, in California the "Daily 3" requires that the player select a three-digit number; if it agrees with the random number chosen, the player receives a $500 payoff. Clearly there is 1 chance in 1000 of being correct and at $1 a ticket, the expected value is $0.50.

\(^{15}\) The term subjective arises from the fact that the probability \( \mu \) is inferred for each decision maker from his or her choices; it is not an objective probability.
an interval scale utility function $U$ over uncertain alternatives,\textsuperscript{16} including the consequences, such that the preference order over alternatives is preserved by

$$U(g) = \sum_{i=1}^{n} S(E_i) U(g_i).$$

This is the expectation of $U$ of the consequences relative to the subjective probability measure $S$. The first fully complete axiomatization of such a representation was by Savage (1954). There have been many subsequent versions, some of which are summarized in Fishburn (1970, 1982, 1988) and Wakker (1989).

2. Who Knows the SEU Formula?

Note the order of information underlying the SEU formulation, Eq. (56): The patterns of preferences are the given empirical information; if they exhibit certain properties (formulated below), they determine the existence of the numerical representation of Eq. (56). It is not assumed that the representation drives the preferences. The representation is a creature of the theorist, and there is no imputation whatsoever that people know $U$ and $S$ and carry out, consciously or otherwise, the arithmetic computations involved in Eq. (56). These remarks are true in the same sense that the differential equation describing the flight path of a ballistic missile is a creation of the physicist, not of the missile.

The mechanisms underlying the observed process—decision or motion—simply are not dealt with in such a theory. Many cognitive psychologists are uncomfortable with such a purely phenomenological approach and feel a need to postulate hypothetical information processing mechanisms to account for what is going on. Busemeyer and Townsend (1993) illustrate such theorizing, and certainly, to the degree that psychology can be reduced to biology, such mechanisms will have to be discovered. At present there is a wide gulf between the mechanisms of cognitive psychologists and biologists.

It should be remarked that throughout the chapter the causal relation between behavior and representation is that the latter, which is for the convenience of the scientist, derives from the former and is not assumed to be a behavioral mechanism. We did not bring up this fact earlier mainly because there has not been much tendency to invert the causal order until one comes to decision theory.

\textsuperscript{16} This use of the symbol $U$ as a function is very different from its earlier use as the second component of a conjoint structure. To be consistent we should use $\phi$, but it is fairly common practice to use $U$ for utility functions.
3. Necessary Properties Underlying SEU

The most basic, and controversial, property underlying SEU and any other representation that says there is an order-preserving numerical representation is that $\succeq$ is a weak order. Although, many accept this as a basic tenet of rationality, others question it both conceptually and empirically. For a thorough summary of the issues and an extensive list of references, see van Acker (1990).

Aside from $\succeq$ being a weak order, the most important necessary properties leading to Eq. (56) are two forms of monotonicity, which can be described informally as follows. Consequence monotonicity means that if any $g_i$ is replaced by a more preferred $g'_i$, with all other consequences and the events fixed, the resulting $g'$ is preferred to $g$. Event monotonicity means that if among the consequences of $g$, $g_1$ is the most preferred and $g_n$ is the least and if $E_k$ is augmented at the expense of $E_m$, then the modified uncertain alternative will be preferred to the original one.

A third property arises from the linear nature of Eq. (56). It is most easily stated for the case in which the chance events are characterized in terms of given probabilities and the representation has the simplifying feature that $S(p_i) = p_i$. In this case we speak of the alternatives as lotteries and the representation obtained from Eq. (56) with $S(E)$ replaced by $p$, as expected utility (EU).

Suppose $g$ and $h$ are lotteries from which a new lottery $(g, p; h, 1 - p)$ is composed. The interpretation is that with probability $p$ one gets to play lottery $g$, and with probability $1 - p$ one gets to play $h$. Then, chance picks one of $g$ and $h$, which is then played independently of the preceding chance decision. When $g$ is run, the consequence $g_i$ occurs with probability $p_i, i = 1, \ldots, n$, and when $h$ is run, the consequence $h_i$ occurs with probability $q_i, i = 1, \ldots, m$. Assuming EU, we see that

$$U(g, p; h, 1 - p) = pU(g) + (1 - p)U(h)$$

$$= p \sum_{i=1}^{n} U(g_i)p_i + (1 - p) \sum_{j=1}^{m} U(h_j)q_j$$

$$= U(g_1, p_1; \ldots, g_n, p_n; h_1, q_1; \ldots; h_m, q_m(1 - p)).$$

Thus, according to the EU representation,

$$(g, p; h, 1 - p) \sim [g_1, p_1; \ldots, g_n, p_n; h_1, q_1; \ldots; h_m, q_m(1 - p)].$$

This property is known as reduction of compound lotteries. Combining consequence monotonicity with the (often implicit) reduction of compound
gambling is known among economists as independence. The use of the reduction-of-compound-gambling principle is implicit when, for example, one assumes, as is common in economics, that the lotteries can be modeled as random variables, in which case Eq. (57) is actually an equality because no distinction is made among various alternative realizations of a random variable.

For uncertain alternatives, a principle, similar in spirit to the reduction of compound lotteries, reads as follows: If two alternatives are identical except for the sequence in which certain events are realized, then the decision maker treats them as equivalent. These are called accounting equivalences (see, e.g., Luce, 1990b). When all conceivable equivalences hold, we speak of universal accounting. Consider the following important specific equivalence. Let \( aO_E b \) denote that \( a \) is the consequence if \( E \) occurs and \( b \) otherwise. Then we say event commutativity holds if

\[
(aO_E b)O_D a \sim (aO_D b)O_E b.
\]

The left term is interpreted to mean that two independent experiments are run and \( a \) is received only if event \( D \) occurs in the first and \( E \) in the second. Otherwise, the consequence is \( b \). The right term is identical except that \( E \) must occur in the first and \( D \) in the second.

Consequence monotonicity and the reduction of compound lotteries are necessary for EU, and they go a long way toward justifying the representation. Similarly, consequence and event monotonicity and universal accounting equivalences are necessary for SEU and they, too, go a long way toward justifying SEU. For this reason, they have received considerable empirical attention.

4. Empirical Violations of Necessary Properties

Perhaps the most basic assumption of these decision models is that preferences are context independent. It is implicitly assumed whenever we attach a utility to an alternative without regard to the set of alternatives from which it might be chosen. To the extent this is wrong, the measurement enterprise, as usually cast, is misguided. MacCrimmon, Stanburg, and Wehrung (1980) have presented very compelling evidence against context independence. They created two sets of lotteries, each with four binary lotteries plus a fixed sum, \( s \). The sum \( s \) and one lottery, \( l \), were common to both sets. Medium level executives, at a business school for midcareer training, were asked (among other things) to rank order each set by preference. A substantial fraction ordered \( s \) and \( l \) differently in the two sets.

---

17 The word independence has many different but related meanings in these areas, so care is required to keep straight which one is intended.
The next most basic assumption to the utility approach is transitivity of preference. To the degree failures have been established, they appear to derive from other considerations. Context effects are surely one source. A second is demonstrated by the famed preference reversal phenomena in which lottery \( g \) is chosen over \( h \) but, when asked to assign monetary evaluations, the subject assigns less value to \( g \) than to \( h \) (Lichtenstein & Slovic, 1971; Luce, 1992b, for a list of references; Slovic & Lichtenstein, 1983). This intransitivity probably reflects a deep inconsistency between judged and choice certainty equivalents rather than being a genuine intransitivity.

Long before intransitivity or context effects were seriously examined, independence and event monotonicity were cast in serious doubt by, respectively, Allais (1953; see Allais & Hagen, 1979) and Ellsberg (1961), who both formulated thought experiments in which reasonable people violate these conditions. These are described in detail in various sources including Luce and Raiffa (1957/1989) and Fishburn (1970). Subsequent empirical work has repeatedly confirmed these results; see Luce (1992b) and Schoemaker (1982, 1990).

The major consequence for theory arising from the failure of event monotonicity is that Eq. (56) can still hold, but only if \( S \) is a weight that is not a probability. In particular, additivity, that is, for disjoint events \( D, E \),

\[
S(D \cup E) = S(D) + S(E)
\]

—although true of probability—cannot hold if event monotonicity is violated. This has led to the development of models leading to the representation of Eq. (56), but with \( S \) a nonadditive weight, not a probability.

The failure of independence is less clear-cut in its significance: Is the difficulty with consequence monotonicity, with the reduction of compound gambles, or both? As Luce (1992b) discussed, there has been an unwarranted tendency to attribute it to monotonicity. This has determined the direction most (economist) authors have taken in trying to modify the theory to make it more descriptive.

Data on the issue have now been gathered. Kahneman and Tversky (1979) reported studies, based on fairly hypothetical judgments, of both independence and monotonicity, with the former rejected and the latter sustained. Several studies (Birnbaum, 1992; Birnbaum, Coffey, Mellers, & Weiss, 1992; Mellers, Weiss, & Birnbaum, 1992) involving judgments of certainty equivalents have shown what seem to be systematic violations of monotonicity. Figure 11 illustrates a sample data plot. In an experimental setting, von Winterfeldt, Chung, Luce, and Cho (1997), using both judgments and a choice procedure, questioned that conclusion, especially when choices rather than judgments are involved.\(^{18}\) They also argued that even

\(^{18}\) This distinction is less clear than it might seem. Many methods exist for which the classification is obscure, but the studies cited used methods that lie at the ends of the continuum.
Evidence is accumulating that judged and choice-determined CEs of gambles simply are not in general the same (Bostic, Herrnstein, & Luce, 1990; Mellers, Chang, Birnbaum, & Ordóñez, 1992; Tversky, Slovic, & Kahneman, 1990). This difference is found even when the experimenter explains what a choice certainty equivalent is and asks subjects to report them directly. Evidence for the difference is presented in Figure 12. A question of some interest is whether, throughout psychophysics, judged indifferences such as curves of equal brightness, fail to predict accurately comparisons between pairs of stimuli. Surprisingly, we know of no systematic study of these matters; it has simply been taken for granted that they should be the same.

Among decision theorists, the most common view is that, to the degree a difference exists, choices are the more basic, and most theoretical approaches have accepted that. One exception is Luce, Mellers, and Chang (1993) who have shown that the preceding data anomalies, including Figure 12, are readily accounted for by assuming that certainty equivalents are basic and the choices are derived from them somewhat indirectly by establishing a reference level that is determined by the choice context, recoding alternatives as gains and losses relative to the reference level, and then using a sign-dependent utility model of a type discussed in section VI.D.2. Indeed, in that section we take up a variety of generalized utility models.

FIGURE 11 Certainty equivalents for binary gambles \((x, p; \$96)\) versus \(1 - p\) for two values of \(x\), \$0 and \$24. Note the nonmonotonicity for the two larger values of \(1 - p\). From Figure 1 of "Violations of Monotonicity and Contextual Effects in Choice-Based Certainty Equivalents," by M. H. Birnbaum, 1992, *Psychological Science*, 3, p. 312. Reprinted with the permission of Cambridge University Press.
1 The Representational Measurement Approach to Problems

OBSERVED

<table>
<thead>
<tr>
<th>OBSERVED</th>
<th>OBSERVED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Choice Proportions</strong></td>
<td><strong>Selling Prices</strong></td>
</tr>
<tr>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>5.40</td>
<td>5.40</td>
</tr>
<tr>
<td>9.70</td>
<td>9.70</td>
</tr>
<tr>
<td>17.50</td>
<td>17.50</td>
</tr>
<tr>
<td>31.50</td>
<td>31.50</td>
</tr>
<tr>
<td>56.70</td>
<td>56.70</td>
</tr>
</tbody>
</table>

FIGURE 12  Rank orders of 36 gambles established from choices and from selling prices (a form of certainty equivalent). The stimuli on the negative diagonal are approximately equal in expected value. Note the sharply different patterns. Adapted from Figure 5 of "Is the Choice Correct Primitive? On Using Certainty Equivalents and Reference Levels to Predict Choices among Gambles," by R. D. Luce, B. Mellers, and S. J. Chang, 1992, Journal of Risk and Uncertainty, 6, p. 133. Reprinted with permission.

D. Functional Measurement

Anderson (1981, 1982, 1991a, b, c) has provided detailed and comprehensive summaries, along with numerous applications to a wide range of psychological phenomena—including psychophysical, personality, and utility judgments—of a method that he and others using the same approach call functional measurement, presumably with the ambiguity intentional. The method begins with a particular experimental procedure and uses, primarily, three types of representations: additive, multiplicative, and averaging. These are described as "psychological laws" relating how the independent variables influence the dependent one.

Stimuli are ordered n-tuples of (often discrete) factors, where n usually varies within an experiment. This differs from conjoint measurement in which the number of factors, n, is fixed. For example, a person may be described along various subsets of several dimensions, such as physical attractiveness, morality, honesty, industry, and so on. Subjects are requested to assign ratings (from a prescribed rating scale) to stimuli that are varied according to some factorial design on the factors. The assigned ratings are viewed as constituting a psychophysical law relating measures of the stimulus to subjective contributions.

Then, assuming that one of the three representations—additive, multiplicative, or averaging—describes the data (usually without any nonlinear transformation of them), Anderson developed computational schemes for estimating the parameters of the representation and for evaluating goodness of fit.
As a simple example, he readily distinguished the additive from averaging representation as follows. Suppose $A_1$ and $A_2$ are two stimulus factors and that $a_1$ and $a_2$ are both desirable attributes but with $a_1$ more desirable than $a_2$. Thus, in an additive representation,

$$\phi_1(a_1) < \phi_1(a_1) + \phi_2(a_2) = \phi(a_1, a_2),$$

whereas in an averaging one

$$\phi_1(a_1) = w\phi_1(a_1) + (1 - w)\phi_1(a_1) > w\phi_1(a_1) + (1 - w)\phi_2(a_2) = \phi(a_1, a_2).$$

This observable distinction generalizes to more than two factors.

Much judgmental data in which the number of factors is varied favors the averaging model. For example, data on person perception make clear, as seems plausible, that a person who is described only as "brilliant" is judged more desirable than one who is described as both "brilliant and somewhat friendly." Anderson's books and papers are replete with examples and experimental detail.

VI. SCALE TYPE, NONADDITIVITY, AND INVARIANCE

In the earlier sections we encountered two apparently unrelated, unresolved issues—the possible levels of nonuniqueness, called scale types, and the existence of nonadditive structures. We examine these issues now. As we shall see, a close relation exists between them and another topic, invariance, only briefly mentioned thus far.

A. Symmetry and Scale Type

1. Classification of Scale Types

As we have noted, numerical representations of a qualitative structure usually are not unique. The nonuniqueness is characterized in what are called uniqueness theorems. We have already encountered three scale types of increasing strength: ordinal, interval, and ratio. The reader may have noted that we said nothing about the uniqueness of threshold structures (section IV.A.3). This is because no concise characterization exists.

S. S. Stevens (1946, 1951), a famed psychophysicist, first commented on the ubiquity of these three types of scales. In a transatlantic debate with members of a commission of the British Association for the Advancement of Science, Stevens argued that what is crucial in measurement is not, as was claimed by the British physicists and philosophers of science, extensive
structures as such but rather structures of any type that lead to a numerical representation of either the interval or, better yet, ratio level.

2. The Automorphism Formulation

More than 30 years later, Narens (1981a, b) posed and formulated the following questions in an answerable fashion: Why these scale types? Are there others? The key to approaching the problem is to describe at the qualitative level what gives rise to the nonuniqueness. It is crucial to note that at its source are the structure-preserving transformations of the structure onto itself—the so-called automorphisms of the structure. These automorphisms describe the symmetries of the structure in the sense that everything appears to be the same before and after the mapping. The gist of the uniqueness theorems really is to tell us about the symmetries of the structure. Thus, the symmetries in the ratio case form a one-parameter family; in the interval case they form a two-parameter family; and in the ordinal case they form a countable family.

Moreover, Narens observed that these three families of automorphisms are all homogeneous: each point of the structure is, structurally, exactly like every other point in the sense that, given any two points, some automorphism maps the one into the other. A second fact of the ratio and interval cases, but not of the ordinal cases, is that when the values of an automorphism are specified at N points (where N is 1 in the ratio case and 2 in the interval one), then it is specified completely. This he called finite uniqueness. An ordinal structure is not finitely unique; it requires countably many values to specify a particular automorphism.

Narens attacked following the question: For the class of ordered structures that are finitely unique and homogeneous and that have representations on the real numbers, what automorphism groups can arise? He developed partial answers and Alper (1987) completed the program. Such structures have automorphism groups of either ratio or interval type or something in between the two. Examples of the latter kind are the sets of numerical transformations \( x \to k^nx + s \), where \( x \) is any real number, \( k \) is a fixed positive number, \( n \) varies over all of the integers, positive and

19 Ratio is better in that it admits far more structures than does the interval form. We will see this when we compare Eqs. (59) and (60). Ratio is stronger than interval in having one less degree of freedom, but it is weaker in the sense of admitting more structures.

20 The term automorphism means "self-isomorphism." Put another way, an automorphism is an isomorphic representation of the structure onto itself.

21 The set of automorphisms forms a mathematical group under the operation \( \ast \) of function composition, which is associative and has an inverse relative to the identity automorphism (see footnote 2).

22 Alper also gave a (very complex) characterization of the automorphism groups of structures that are finitely unique but not homogeneous.
negative, and $s$ is any real number; these are 1-point homogeneous and 2-point unique.

The key idea in Alper’s proof is this. An automorphism is called a translation if either it is the identity or no point of the structure stays fixed under the automorphism. For example, all the automorphisms of a ratio-scale structure are translations, but only some are translations in the case of interval scales. The ordering in the structure induces an ordering on the translations: If $\tau$ and $\sigma$ are two translations, define $\tau \preceq \sigma$ if and only if $\tau(a) \preceq \sigma(a)$ for all $a \in A$. The difficult parts of the proof are, surprisingly, in showing that the composition of two translations, $\tau \circ \sigma(a) = \tau(\sigma(a))$, is also a translation, so $\ast$ is an operation on the translations, and that this group of translations is itself homogeneous. The ordered group of translations is also shown to be Archimedean, so by Hölder’s theorem (section IV.C.3) it can be represented isomorphically in $(\mathbb{R}_+, \preceq, \ast)$. Moreover, using the homogeneity of the translations one can map the structure itself isomorphically into the translations and thus into $(\mathbb{R}_+, \preceq, \ast)$. This is a numerical representation in which the translations appear as multiplication by positive constants. In that representation, any other automorphism is proved to be a power transformation $x \mapsto sx^r$, $r > 0, s > 0$.

The upshot of the Narens-Alper development is that as long as one deals with homogeneous structures that are finitely unique and have a representation onto the real numbers, none lie between the interval and the ordinal-like cases. Between the ratio and interval cases are other possibilities. We know of mathematical examples of these intermediate cases that exhibit certain periodic features (Luce & Narens, 1985), but so far they seem to have played no role in science. When it comes to nonhomogeneous structures, little that is useful can be said in general, but some that are nearly homogeneous are quite important (see section VI.D).

B. Nonadditivity and Scale Type
1. Nonadditive Unit Concatenation Structures

The preceding results about scale type are not only of interest in understanding the possibilities of measurement, but they lead to a far more complete understanding of specific systems. For example, suppose $\mathcal{K} = (A, \preceq, \ast)$ is a homogeneous and finitely unique concatenation structure (section 23).

23 Structures with singular points that are fixed under all automorphisms—minimum, maximum, or zero points—are excluded in Alper’s work. They are taken up in section VI.C.

24 Luce and Alper (in preparation) have shown that the following conditions are necessary and sufficient for such a representation: The structure is homogeneous, any pair of automorphisms cross back and forth only finitely many times, the set of translations is Archimedean, and the remaining automorphisms are Archimedean relative to all automorphisms.
IV.C.2) that is isomorphic to a real concatenation structure $\mathbb{R} = (\mathbb{R}, \approx, \oplus)$. By the Narens-Alper theorem, we may assume that $\mathbb{R}$ has been chosen so that the translations are multiplication by positive constants. From that, Luce and Narens (1985), following more specific results in Cohen and Narens (1979), showed, among other things, that $\oplus$ has a simple numerical form, namely, there is a strictly increasing function $f: \mathbb{R} \rightarrow \mathbb{R}$ that also has two properties: $f(x)/x$ is a strictly decreasing function of $x$ and

$$x \oplus y = yf(x/y).$$

(59)

Such structures are referred to as unit concatenation structures. The familiar extensive case is $f(z) = z + 1$, that is, $x \oplus y = x + y$. Equation (59) shows that unit concatenation structures have a very simple structure and that if one is confronted with data that appear to be nonadditive, one should attempt to estimate $f$.

The function $f$ can be constructed as follows. For each natural number $n$, define $\theta(a, n) = \theta(a, n - 1) \circ a$ and $\theta(a, 1) = a$. This is an inductive definition of one sense of what it means to make $n$ copies of the element $a$ of the structure. These are called $n$-copy operators, and it can be shown that they are in fact translations of the structure. In essence, they act like the equally spaced markers on a ruler, and the isomorphism $\phi$ into the multiplicative reals can be constructed from them exactly as in extensive measurement. Once $\phi$ is constructed, one constructs $f$ as follows: For any positive number $z$, find elements $a$ and $b$ such that $z = \phi(a) / \phi(b)$; then

$$f(z) = \phi(a \circ b) / \phi(b).$$

Note that an empirical check is implicit in this, namely that

$$\frac{\phi(a)}{\phi(b)} = \frac{\phi(c)}{\phi(d)} \quad \text{implies} \quad \frac{\phi(a \circ b)}{\phi(b)} = \frac{\phi(c \circ d)}{\phi(d)}.$$

Because a structural property satisfied by any element of a homogeneous structure is also satisfied by all other elements of the structure, such structures must necessarily satisfy, for all $a \in A$, one of the following: (positive) $a \circ a > a$; (negative) $a \circ a < a$; or (idempotent) $a \circ a = a$. It turns out that only the latter can be an interval scale case, and its representation on $\mathbb{R}$ (not $\mathbb{R}^+$) has the following simple rank-dependent form: For some constants $c, d \in (0, 1)$ and all $x, y \in \mathbb{R}$,

$$x \oplus y = \begin{cases} cx + (1 - c)y, & \text{if } x \geq y \\ dx + (1 - d)y, & \text{if } x < y. \end{cases}$$

(60)

25 Recall that the structure is not associative, so a binary operation can be grouped in a large number of ways to form $n$ copies of a single element. We have simply selected one of these, a so-called right-branching one.
This was the form mentioned in section V.C.4. It was first suggested in a psychological application by Birnbaum, Parducci, and Gifford (1971) and used in later papers (Birnbaum, 1974; Birnbaum & Stegner, 1979). As we shall see, it subsequently was rediscovered independently by economists and has been fairly widely applied in utility theory (see Quiggin, 1993, and Wakker, 1989).

2. Homogeneous Conjoint Structures

Without going into much detail, the construction outlined in section IV.C.1 for going from a conjoint to a concatenation structure did not depend on double cancellation being satisfied, so one can use that definition for more general conjoint structures. Moreover, the concept of homogeneity is easily formulated for the general case, and it forces homogeneity to hold in the induced concatenation structure. This reduction makes possible the use of the representation Eq. (59) to find a somewhat similar one for these nonadditive conjoint structures. The details are presented in Luce et al. (1990). Therefore, once again we have a whole shelf of nonadditive representations of ratio and interval types. Of these, only the rank-dependent cases have thus far found applications, but we anticipate their more widespread use once psychologists become aware of these comparatively simple possibilities.

3. Combining Concatenation and Conjoint Structures

Recall that we discussed additive conjoint structures with extensive structures on the components as a model of many simple physical laws and as the basis of the units of physics (section IV.C.7). This result has been generalized to unit concatenation structures, Eq. (59). Suppose a (not necessarily additive) conjoint structure has unit concatenation structures on its components and that the distribution property, Eq. (38), holds. Then one can show that the conjoint structure must in fact be additive and that the representation is that of Eq. (39). Indeed, if the components are endowed with ratio scale structures of any type, not necessarily concatenation structures, then a suitable generalization of distribution is known so that Eq. (39) continues to hold (Luce, 1987).

The upshot of these findings is that it is possible, in principle, to extend the structure of physical quantities to incorporate all sorts of measurements in addition to extensive ones without disturbing the pattern of units being products of powers of base units. Such an extension has yet to be carried out, but we now know that it is not precluded just because an attribute fails to be additive. For additional detail, see Luce et al. (1990, pp. 124--126).
C. Structures with Discrete Singular Points

1. Singular Points

As was remarked earlier, the class of nonhomogeneous structures is very diverse and ill understood. However, one class is quite fully understood, and it plays a significant role in measurement. We discuss that class next.

A point of a structure is called singular if it stays fixed under every automorphism of the structure. The many familiar examples include any minimum point, such as zero length or zero mass; any maximum point, such as the velocity of light; and certain interior points, such as the status quo in utility measurement. Still another example arises in a class of preference models proposed by Coombs and summarized in his 1964 book (see also Coombs, 1975). He postulated that an individual’s preferences arise from a comparison of the relative “distances” between each alternative and that individual’s ideal point on the attribute for which preference is being expressed. Clearly, such an ideal point plays a distinctive role, namely, it is the zero point of dis-preference for that person. Coombs developed algorithms that use the data from a number of subjects to infer simultaneously the location of objects and ideal points in the space of preferences. However, the mathematical theory was never very fully developed. In contrast, the role of the status quo in utility theory is better analyzed.

2. Homogeneity Between Discrete Singular Points

Singular points have properties that render them unlike any other point in the structure; indeed, they keep the structure from being homogeneous. However, if a finite number of singular points exist, as is true in the applications just mentioned, the structure can be homogeneous between adjacent points. That is, if \( a \) and \( b \) are two points of the structure not separated by a singular point, then some automorphism of the structure takes \( a \) into \( b \). Furthermore, if the structure has a generalized monotonic operation, and if it is finitely unique, it can be shown (Luce, 1992a) that there are at most three singular points: a minimum, a maximum, and an interior one. Moreover they exhibit systematic properties. One then uses the results on unit structures to derive a numerical representation of this class of structures. Results about such structures underlie some of the developments in the next subsection.

---

26 As a rough analogy, homogeneous and nonhomogeneous stand in the same relation as do linear and nonlinear equations: The former is highly special, and the latter highly diverse.

27 This is a function of two or more variables that is monotonic in each. One must be quite careful in formulating the exact meaning of monotonicity at minima and maxima.
3. Generalized Linear Utility Models

A growing literature is focused on exploring ways to modify the EU and SEU models (sections V.C.1 and 2) so as to accommodate some of the anomalies described in section V.C.3.28 One class of models, which includes Kahneman and Tversky's (1979; Tversky & Kahneman, 1992) widely cited representation called prospect theory, draws on generalized concatenation structures with singular points, identifying the status quo as a singular point. The resulting representation modifies SEU, Eq. (56), to the extent of making \( S(E) \) depend on one or both of two things beside the event \( E_0 \), namely, the sign of the corresponding consequence \( g_i \)—that is, whether it is a gain or loss relative to the status quo—and also the rank-order position of \( g_i \) among all of the consequences, \( g_1, \ldots, g_n \), that might arise from the gamble \( g \). These models go under several names, including rank- and sign-dependent utility (RSDU) and cumulative prospect theory. In the binary case, such models imply event commutativity Eq. (58), but none of the more complex accounting equivalences that hold for SEU (Luce & von Winterfeldt, 1994).

Measurement axiomatizations of the most general RSDU are given by Luce and Fishburn (1991, 1995) and Wakker and Tversky (1993). The former is unusual in this literature because it introduces a primitive beyond the preference ordering among gambles, namely, the idea of the joint receipt of two things. Therefore, if \( g \) and \( h \) are gambles, such as two tickets in different state lotteries or stock certificates in two corporations, a person may receive (e.g., as a gift) both of them, which is denoted \( g \oplus h \). This operation plays two useful features in the theory. One, which Tversky and Kahneman (1986) called segregation and invoked in pre-editing gambles, states that if \( g \) is a gamble and \( s \) is a certain outcome with the consequences being either all gains or all losses, \( g \oplus s \) is treated as the same as the gamble \( g' \) which is obtained by replacing each \( g_i \) by \( g_i \oplus s \). This appears to be completely rational. The second feature, called decomposition, formulates the single nonrationality of the theory: Let \( g \) be a gamble having both gains and losses, let \( g^+ \) denote the gamble resulting from \( g \) by replacing all of the losses by the status quo, and \( g^- \) that by replacing the gains by the status quo. Then,

\[
g - g^+ \oplus g^-,
\]

where the two gambles on the right are realized independently. This is, in reality, a formal assertion of what is involved in many cost-benefit analyses, the two components of which are often carried out by independent groups of analysts and their results are combined to give an overall evaluation of the

28 Most of the current generalizations exhibit neither context effects, as such, nor intransitivities.
situation. In fact, Slovic and Lichtenstein (1968), in a study with other goals, tested decomposition in a laboratory setting and found it sustained. More recently Cho, Luce, and von Winterfeldt (1994) carried out a somewhat more focused study, again finding good support for the segregation and decomposition assumptions.

Within the domain of lotteries,29 economists have considered other quite different representations. For example, Chew and Epstein (1989) and Chew, Epstein, and Segal (1991) have explored a class of representations called quadratic utility that takes the form

\[ U(g) = \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(g_i, g_j)p_ip_j. \]  

A weakened form of independence is key to this representation. It is called mixture symmetry and is stated as follows: If \( g \sim h \), then for each \( \alpha \in (0, 1) \), there exists \( \beta \in (\bar{4}, 1) \) such that

\[ (g, \alpha; h, 1 - \alpha) \sim (h, \beta; g, 1 - \beta). \]  

Equation (63) and consequence monotonicity together with assumptions about the richness and continuity of the set of lotteries imply that \( \beta = 1 - \alpha \) and that Eq. (62) is order preserving. We are unaware of any attempts to study this structure empirically.

D. Invariance and Homogeneity

1. The General Idea

A very general scientific meta-principle asserts that when formulating scientific propositions one should be very careful to specify the domain within which the proposition is alleged to hold. The proposition must then be formulated in terms of the primitives and defining properties of that domain. When the domain is rich in automorphisms, as in homogeneous cases or in the special singular cases just discussed, this means that the proposition must remain invariant with respect to the automorphisms, just as is true—by definition—of the primitives of the domain.

2. An Example: Bisection

Let \( \mathcal{A} = (A, \succeq, \circ) \) be an extensive structure such as the physical intensity of monochromatic lights. It has a representation \( \phi \) that maps it into \( \mathcal{R} = (\mathbb{R}_+, \succeq, +) \) and the automorphisms (translations) become multiplication by positive constants. Now, suppose a bisection experiment is performed such that when stimuli \( x = \phi(a) \) and \( y = \phi(b) \) are presented, the subject reports

29 Money gambles with known probabilities for the consequences.
the stimulus \( z = \phi(t) \) to be the bisection point of \( x \) and \( y \). We may think of this as an operation defined in \( \mathbb{R} \), namely, \( z = x \oplus y \). If this operation is expressible within the structure \( \mathcal{A} \), then invariance requires that for real \( r > 0 \),

\[
r(x \oplus y) = rx \oplus ry.
\]

A numerical equation of this type is said to be homogeneous of degree 1, which is a classical concept of homogeneity. It is clearly very closely related to the idea of the structure being homogeneous. In section VII we will see how equations of homogeneity of degree different from 1 also arise.

Plateau (1872) conducted a bisection experiment using gray patches, and his data supported the idea that there is a "subjective" transformation \( U \) of physical brightness that maps the bisection operation into a simple average, that is,

\[
U(x \oplus y) = \frac{U(x) + U(y)}{2}.
\]

If we put Eqs. (64) and (65) together, we obtain the following constraint on the function \( U \): For all \( x, y, z \in \mathbb{R}_+ \),

\[
U[RU^{-1} \left( \frac{U(x) + U(y)}{2} \right)] = \frac{U(rx) + U(ry)}{2}.
\]

An equation of this type in which a function is constrained by its values at several different points is called a functional equation (Aczél, 1966, 1987). Applying the invariance principle to numerical representations quite typically leads to functional equations.

In this case, under the assumption that \( U \) is strictly increasing, it can be shown to have one of two forms:

\[
U(x) = k \log x + c \quad \text{or} \quad U(x) = ax^b + d.
\]

These are of interest because they correspond to two of the major proposals for the form of subjective intensity as a function of physical intensity. The former was first strongly argued for by Fechner, and the latter, by Stevens.

Falmagne (1985, chap. 12) summarized other, somewhat similar, invariance arguments that lead to functional equations. We discuss related, but conceptually quite distinct, arguments in section VII.

3. Invariance in Geometry and Physics.

In 1872, the German mathematician F. Klein, in his famous Erlangen address, argued that within an axiomatic formulation of geometry, the only entities that should be called "geometric" are those that are invariant under

\( ^{30} \) The degree of the homogeneity refers to the exponent of the left-hand \( r \), which as written is 1.
the automorphisms of the geometry (for a recent appraisal, see Narens, 1988). Klein used this to good effect; however, a number of geometries subsequently arose that were not homogeneous and, indeed, in which the only automorphism was the trivial one, the identity. Invariance in such cases establishes no restrictions whatsoever. This illustrates an important point—namely, that invariance under automorphisms is a necessary condition for a concept to be formulated in terms of the primitives of a system, but it is by no means a sufficient condition.

During the 19th century, physicists used, informally at first, invariance arguments (in the form of dimensional consistency) to ensure that proposed laws were consistent with the variables involved. Eventually this came to be formulated as the method of dimensional analysis in which numerical laws are required to be dimensionally homogeneous of degree 1 (dimensional invariance). Subsequently, this method was given a formal axiomatic analysis (Krantz et al., 1971, chap. 10; Luce et al., 1990, chap. 22), which showed that dimensional invariance is, in fact, just automorphism invariance. Again, it is only a necessary condition on a physical law, but in practice it is often a very restrictive one. Dzhafarov (1995) presented an alternative view that is, perhaps, closer to traditional physical presentations. Nontrivial examples of dimensional analysis can be found in Palacios (1964), Sedov (1959), and Schepartz (1980). For ways to weaken the condition, see section VII.

4. Invariance in Measurement Theory and Psychology

In attempting to deal with general structures of the type previously discussed, measurement theorists became very interested in questions of invariance, for which they invented a new term. A proposition formulated in terms of the primitives of a system is called meaningful only if it is invariant under the automorphisms of the structure. Being meaningful says nothing, one way or the other, about the truth of the proposition in question, although meaningfulness can be recast in terms of truth as follows: A proposition is meaningful if it is either true in every representation (within the scale type) of the structure or false in every one. Being meaningless (not meaningful) is not an absolute concept; it is entirely relative to the system in question, and something that is meaningless in one system may become meaningful in a more complete one. As noted earlier, the concept has bite only when there are nontrivial automorphisms. Indeed, in a very deep and thorough analysis of the concept of meaningfulness, Narens (in preparation) has shown that it is equivalent to invariance only for homogeneous structures.

In addition to meaningfulness arguments leading to various psychophysical equations, such as Eqs. (65) through (67), some psychologists, beginning with Stevens (1951), have been involved in a contentious controversy about applying invariance principles to statistical propositions. We do
not attempt to recapitulate the details. Suffice it to say that when a statistical proposition is cast in terms of the primitives of a system, it seems reasonable to require that it be true (for false) in every representation of the system. Thus, in the ordinal case it is meaningless to say (without further specification of the representation) that the mean of one group of subjects is less than the mean of another because the truth is not invariant under strictly increasing mappings of the values. In contrast, comparison of the medians is invariant. A list of relevant references can be found in Luce et al. (1990). See also Michell (1986) and Townsend and Ashby (1984).

VII. MATCHING AND GENERAL HOMOGENEOUS EQUATIONS

A. In Physics

Many laws of physics do not derive from the laws that relate basic physical measures but are, nonetheless, expressed in terms of these measures. This section describes two such cases, which are handled differently.

1. Physically Similar Systems

Consider a spring. If one holds the ambient conditions fixed and applies different forces to the spring, one finds, within the range of forces that do not destroy the spring, that its change in length, $\Delta l$, is proportional to the force, $F$, applied, $\Delta l = kF$. This is called Hooke's law. Note that such a law, as stated, is not invariant under automorphisms of the underlying measurements. This is obvious because $\Delta l$ has the dimension of length, whereas force involves mass, length, and time. This law is expressible in terms of the usual physical measures, but it is not derivable from the underlying laws of the measurement structure.

A law of this type can be recast in invariant form by the following device. The constant $k$, called the spring constant, is thought of as characterizing a property of the spring. It varies with the shape and materials of the spring and has units that make the law dimensionally invariant, namely, $[L][T]^2/[M]$, where $[L]$ denotes the unit of length, $[T]$ of time, and $[M]$ of mass. This is called a dimensional constant. Such constants play a significant role in physics and can, in general, be ascertained from the constants in the differential equations characterizing the fundamental physical relations underlying the law when these equations are known. The set of entities characterized in this fashion, such as all springs, are called physically similar systems.

2. Noninvariant Laws

In addition to invariant laws or laws that can be made invariant by the inclusion of dimensional constants that are thought to be characteristic of the system involved, there are more complex laws of which the following is
an example taken from rheology (material science) called a Nutting law (Palacios, 1964, p. 45). For solids not satisfying Hooke's law (see the preceding), the form of the relationship is \( d = k(F/A)\beta t^{\gamma} \), where \( d = \Delta l/l \) is the deformation and \( F \) the applied force. New are the area \( A \) to which \( F \) is applied, the time duration \( t \) of the application, and the two exponents \( \beta \) and \( \gamma \), both of which depend on the particular material in question. Thus, the dimension of \( k \) must be \( [L]^p[m]^{-p}T^{\beta - \gamma} \), which of course varies depending on the values of \( \beta \) and \( \gamma \).

To understand the difficulty here, consider the simplest such case, namely \( \gamma = kx^p \), where \( x \) and \( y \) are both measured as (usually, distinct) ratio scales. Thus, the units of \( k \) must be \( [y]/[x]^p \). There is no problem so long as all systems governed by the law have the same exponent \( \beta \); the situation becomes quite troublesome, however, if not only the numerical value of \( k \) depends on the system but, because of changes in the value of \( \beta \), the units of \( k \) also depend on the particular system involved. Such laws, which are homogeneous of degree \( \beta \), cannot be made homogeneous of degree 1 by introducing a dimensional constant with fixed dimensions. In typical psychology examples, several of which follow, the value of \( \beta \) appears to vary among individuals just as it varies with the substance in rheology. Both Falmagne and Narens (1983) and Dzhafarov (1995) have discussed different approaches to such problems.

There is a sense, however, in which invariance is still nicely maintained. The ratio scale transformation (translation) \( r \) on the dimension \( x \) is taken into a ratio scale transformation \( r^p \) on the dimension \( y \). Put another way, the law is compatible with the automorphism structures of dimensions \( x \) and \( y \) even if it is not invariant with respect to the automorphisms (Luce, 1990a). Such homogeneous laws are very useful in psychophysics because they narrow down to a limited number of possibilities the mathematical form of the laws (see Falmagne, 1985, chap. 12).

### B. Psychological Identity

Some psychophysical laws formulate conditions under which two stimuli are perceived as the same. We explore two illustrations.

1. *Weber-Type Laws*

Consider a physical continuum of intensity that can be modeled as an extensive structure \( (A, \succsim, \emptyset) \) with the ratio scale (physical) representation \( \phi \) onto \( (\mathbb{R}, +, <) \). In addition, suppose there is a psychological ordering \( >_\Psi \) on \( A \) that arises from a discrimination experiment where, for \( a, b \in A \), \( b >_\Psi a \)

---

31 We thank E. N. Dzhafarov (personal communication) for bringing these rheology examples to our attention. See Scott Blair (1969) for a full treatment.

32 The subscript \( \Psi \) is intended as a reminder that this ordering is psychological and quite distinct from \( \succsim \), which is physical.
means that $b$ is perceived as more intense than $a$. In practice, one estimates a psychometric function and uses a probability cutoff to define $>_{\Psi}$; see Eq. (20) of section IV. Not surprisingly, such orderings are usually transitive. However, if we define $b \sim_{\Psi} a$ to mean neither $b >_{\Psi} a$ nor $a >_{\Psi} b$, then in general $\sim_{\Psi}$ is not transitive: The failure to discriminate $a$ from $b$ and $b$ from $c$ does not necessarily imply that $a$ cannot be discriminated from $c$, although that may happen. It is usually assumed that $>_{\Psi}$ satisfies the technical conditions of a semiorder or an interval order (see section IV.A.3), but we do not need to go into those details here.

Narens (1994) proved that one cannot define the structure $(A, \geq, \odot)$ in terms of $(A, >_{\Psi})$; however, one can formulate the latter in terms of the former in the usual way. One defines $T(a)$ to be the smallest (inf) $b$ such that $b >_{\Psi} a$. Then $T$ establishes a law that maps $A$ into $A$, namely, the upper threshold function. Typically, this is converted into a statement involving increments. Define $\Delta(a)$ to be the element such that $T(a) = a \odot \Delta(a)$. Auditory psychologists (Jesteadt, Wier, & Green, 1977; McGill & Goldberg, 1968) have provided evidence (see Figure 2) that intensity discrimination of pure tones exhibits the property that the psychologically defined $\Delta(a)$ is compatible with the physics in the sense that for each translation $T$ of the physical structure, there is another translation $\sigma_{\tau}$, dependent on $\tau$, such that

$$\Delta(\tau(a)) = \sigma_{\tau}(\Delta(a)).$$

(68)

When recast as an equivalent statement in terms of the representations, Eq. (68) asserts the existence of constants $\alpha > 0$ and $\beta > 0$ such that

$$\phi(\Delta(a)) = \alpha \phi(a)^{1-\beta},$$

(69)

which again is a homogeneous equation of degree $1 - \beta$ (Luce, 1990a).

The latter formulation is called the near miss to Weber's law because when $\phi$ is the usual extensive measure of sound intensity, $\beta$ is approximately 0.07, which is "close to" $\beta = 0$, the case called Weber's law after the 19th-century German physiologist E. H. Weber. Note that Weber's law itself is special because it is dimensionally invariant, that is, in Eq. (68) $\sigma_{\tau} = \tau$, but the general case of Eq. (68) is not. It is customary to rewrite Weber's law as

$$\frac{\Delta\phi(a)}{\phi(a)} = \frac{\phi(\Delta(a))}{\phi(a)} = \alpha.$$  

(70)

This ratio, called the Weber fraction, is dimensionless and some have argued that, to the extent Weber's law is valid, the fraction $\alpha$ is a revealing parameter of the organism, and, in particular, that it is meaningful to compare Weber fractions across modalities. This common practice has recently been questioned by measurement theorists, as we now elaborate.
2. Narens-Mausfeld Equivalence Principle

Narens (1994) and Narens and Mausfeld (1992) have argued that one must be careful in interpreting the constants in laws like Eq. (69) and (70). They note that the purely psychological assertion about discriminability has been cast in terms of one particular formulation of the qualitative physical structure, whereas there are an infinity of concatenation operations all of which are equally good in the following sense. Two qualitative formulations \( \langle A, \succeq, \odot \rangle \) and \( \langle A, \succeq, * \rangle \) are equivalent if each can be defined in terms of the other. This, of course, means that they share a common set of automorphisms: If \( \tau \) is an automorphism of one, then it is of the other; that is, both \( \tau(a \odot b) = \tau(a) \odot \tau(b) \) and \( \tau(a * b) = \tau(a) * \tau(b) \) hold. Indeed, Narens (1994) has shown that if the former has the ratio scale representation \( \phi \), then the latter must have one that is a power function of \( \phi \). Thus, if the former structure is replaced by the latter, then Eq. (70) is transformed into

\[
\frac{\phi(y^*(a))}{\phi(y(a))} = (1 + \alpha)^y - 1,
\]

where \( y \) is chosen so \( \phi^y \) is additive over \(*\). Thus, the fact that Weber's law holds is independent of which physical primitives are used to describe the domain, and so within one modality one can compare individuals as to their discriminative power. Across modalities, no such comparison makes sense because the constant \( (1 + \alpha)^y - 1 \) is not invariant with the choice of the concatenation operation, which alters the numerical value of \( y \).

If one reformulates the law in terms of \( T(a) = a \odot \Delta(a) \), Weber's law becomes

\[
\frac{T(a)}{\phi(a)} = 1 + \alpha.
\]

Note that this formulation does not explicitly invoke a concatenation operation, except that choosing \( \phi \) rather than \( \phi^y \) does, and so the same strictures of interpretation of the ratios remain.\(^{33}\)

Carrying out a similar restatement of the near miss, Eq. (69) yields

\[
\frac{\phi(y^*(T(a)))}{\phi(y(T(a)))} = (1 + \alpha \phi(y(a))^{-\beta/y})^y.
\]

Here the choice of a concatenation operation clearly affects what one says about the "near-miss" exponent because the value \( \beta/y \) can be anything.

The principle being invoked is that psychologically significant propositions can depend on the physical stimuli involved, but they should not depend on the specific way we have chosen to formulate the physical situation.

\(^{33}\) This remark stands in sharp contrast to Narens' (1994) claim that \( \alpha + 1 \) is meaningful under circumstances when \( \alpha \) is not.
We should be able to replace one description of the physics by an equivalent one without disturbing a psychologically significant proposition.

This principle is being subjected to harsh criticism, the most completely formulated of which came from Dzhafarov (1995) who argued that its wholesale invocation will prove far too restrictive not only in psychology but in physics as well. It simply may be impossible to state psychological laws without reference to a specific formulation of the physics, as appears likely to be the case in the next example.

3. Color Matching

A far more complex and interesting situation arises in color vision. The physical description of an aperture color is simply the intensity distribution over the wave lengths of the visible spectrum. A remarkable empirical conclusion is that there are far fewer color percepts than there are intensity distributions: The latter form an infinite dimensional space that, according to much psychological data and theory, human vision collapses into a much lower dimensional one—under some circumstances to three dimensions. The experimental technique used to support this hypothesis is called *metamer matching* in which a circular display is divided into two half-fields, each with a different intensity distribution. When a subject reports no perceived difference whatsoever in the two distributions, which may be strikingly different physically, they are said to match.

One possible physical description of the stimuli is based on two easily realized operations. Suppose \( a \) and \( b \) denote two intensity distributions over wave length. Then \( a \oplus b \) denotes their sum, which can be achieved by directing two projectors corresponding to \( a \) and \( b \) on the same aperture. For any real \( r > 0 \), \( ra \) denotes the distribution obtained from \( a \) by increasing every amplitude by the same factor \( r \), which can be realized by changing the distance of the projector from the aperture. In terms of this physical structure and the psychological matching relation, denoted \( \sim \), Krantz (1975a, b; Suppes et al., 1989, chap. 15) has formulated axiomatically testable properties of \( \sim \), \( \oplus \), and of their interactions that, if satisfied, result in a three-dimensional vector representation of these matches. Empirical data provide partial, but not full, support for these so-called Grassman laws.

The dimension of the representation is an invariant, but there are infinitely many representations into the vector space of that dimension. A substantial portion of the literature attempts to single out one or another as having special physiological or psychological significance. These issues are described in considerable detail in chapter 15 of Suppes et al. (1989), but as yet they are not fully resolved.

---

34 It is worth noting that sounds also are infinite dimensional, but no such reduction to a finite dimensional perceptual space has been discovered.
To our knowledge no attempt has been made to analyze these results from the perspective of the Narens-Mausfeld principle. It is unclear to us what freedom exists in providing alternative physical formulations in this case.

C. Psychological Equivalence

1. Matching across Modalities

As was discussed in section IV.B.1, psychologists often ask subjects to characterize stimuli that are equivalent on some subjective dimension even though they are perceptually very distinct. Perhaps the simplest cases are the construction of equal-X curves, where X can be any suitable attribute: brightness, loudness, aversiveness, and so on. Beginning in the 1950s, S. S. Stevens (1975) introduced three new methods that went considerably beyond matching within an attribute: magnitude estimation, magnitude production, and cross-modal matching. Here two distinct attributes—a sensory attribute and a numerical attribute in the first two and two sensory attributes in the third—are compared and a "match" is established by the subject. The main instruction to subjects is to preserve subjective ratios. Therefore if \( M \) denotes the matching relation and \( aMs \) and \( bMt \), then the instruction is that stimuli \( a \) and \( b \) from modality \( A \) should stand in the same (usually intensity) subjective ratio as do \( s \) and \( t \) from modality \( S \).

In developing a theory for such matching relations, the heart of the problem is to formulate clearly what it means "to preserve subjective ratios." In addition, of course, one also faces the issue of how to deal with response variability, which is considerable in these methods, but we ignore that here. Basically, there are three measurement-theoretic attempts to provide a theory of subjective ratios.

The first, due to Krantz (1972) and Shepard (1978, 1981), explicitly introduced as a primitive concept the notion of a ratio, formulated plausible axioms, and showed that in terms of standard ratio scale representations of the physical attributes, \( \phi_A \) and \( \phi_S \), the following is true for some unspecified monotonic function \( F \) and constant \( \beta > 0 \):

\[
aMs \text{ and } bMt \text{ if and only if } \frac{\phi_A(b)}{\phi_A(a)} = F \left[ \left( \frac{\phi_S(t)}{\phi_S(s)} \right)^\beta \right]. \quad (74)
\]

Although the power function character is consistent with empirical observations, the existence of the unknown function \( F \) pretty much obviates that relationship.

A second attempt, due to Luce (1990a, and presented as an improved formulation of Luce, 1959b), stated that ratios are captured by translations. In particular, he defined a psychological matching law \( M \) to be translation consistent (with the physical domains \( A \) and \( S \)) if for each translation \( \tau \) of the
domain $A$ there exists a corresponding translation $\sigma$, of the domain $S$ such that for all $a \in A$ and $s \in S$,

$$a Ms \text{ if and only if } \tau(a) M_{\sigma}(s). \tag{75}$$

From this it follows that if $\phi_A$ and $\phi_S$ are ratio scale representations of the two physical domains, then there are constants $\alpha > 0$ and $\beta > 0$ such that

$$a Ms \text{ is equivalent to } \phi_A(a) = \alpha \phi_S(s)^{\beta}. \tag{76}$$

Observe that Eqs. (68) and (69) are special cases of (75) and (76).

The third attempt, due to Narens (1996), is far deeper and more complex than either of the previous attempts. He carefully formulated a plausible model of the internal representation of the stimuli showing how the subject (in magnitude estimation) constructing numerals to produce responses. It is too complex to describe briefly, but any serious student of these methods should study it carefully.

2. Ratios and Differences

Much of the modeling shown in section II was based on functions of differences of subjective sensory scales. Similarly, methods such as bisection and fractionation more generally seem to rest on subjects evaluating differences. By contrast, the discussion of cross-modal matching (and of magnitude estimation and production) emphasizes the preservation of ratios. Torger-son (1961) first questioned whether subjects really have, for most dimensions, independent operations corresponding to differences and to ratios, or whether there is a single operation with two different response rules depending on the instructions given. Michael Birnbaum, our editor, has vigorously pursued this matter.

The key observation is that if there really are two operations, response data requiring ratio judgments cannot be monotonically related to those requiring difference judgments. For example, $3 - 2 < 13 - 10$ but $1.5 = 3/2 > 13/10 = 1.3$. On the other hand, if ratio judgments are found to covary with difference judgments in a monotonic fashion, a reasonable conclusion is that both types of judgments are based on a single underlying operation.

A series of studies in a variety of domains ranging from physical manipulable attributes (such as weight and loudness) to highly subjective ones (such as job prestige) has been interpreted as showing no evidence of non-monotonicity and to provide support for the belief that the basic operation is really one of differences. The work is nicely summarized by Hardin and Birnbaum (1990), where one finds copious references to earlier work.

Hardin and Birnbaum conclude that the data support a single operation that involves subtracting values of a subjective real mapping $s$, and that
depending on the task required of the subject, different response functions are employed for the two judgments, namely,

\[ \text{Response} = f_D[s(a) - s(b)] \text{ for difference judgments} \]

and

\[ \text{Response} = f_R[s(a) - s(b)] \text{ for ratio judgments.} \] (77)

Moreover, the evidence suggests that approximately \( f_R(x) = \exp f_D(x). \)

They do point out that, for a few special modalities, ratios and differences can be distinct. This is true of judgments of length: most people seem to understand reasonably clearly the difference between saying two height ratios are equal and that two differences in length are equal.

The empirical matter of deciding if ratio and difference judgments are or are not monotonically related is not at all an easy one. One is confronted by a data figure such as that reproduced in Figure 13 and told that it represents a single monotonic function. But are the deviations from a smooth monotone curve due to response error, or "noise," or are they small but systematic indications of a failure of monotonicity? It is difficult to be sure in average data such as these. A careful analysis of the data from individual subjects might be more convincing, however; and Birnbaum and Elmasian (1977) carried out such an analysis, concluding that a single operation does give a good account of the data.

![Mean "Ratio" vs. Mean "Difference" Judgment](image-url)

**FIGURE 13** Geometric mean estimates of ratio judgments versus mean difference judgments of the same stimulus pairs of occupations. From Figure 1 of "Malleability of 'Ratio' Judgments of Occupational Prestige," by C. Hardin and M. H. Birnbaum, 1990, *American Journal of Psychology*, 103, p. 6. From *American Journal of Psychology*. Copyright 1990 by the Board of Trustees of the University of Illinois. Used with the permission of the University of Illinois Press.
Assuming that the issue of monotonicity has been settled, there remains the question whether the underlying operation is one of differences or ratios. For, as is well known, we can replace the right-hand terms in Eq. (77) by corresponding expressions involving ratios rather than differences, namely

\[ J_B[s'(a)/s'(b)] \text{ and } J_B[s'(a)/s'(b)], \]

where \( s'(a) = \exp[s(a)], \)
\[ J_D(x) = J_D[\ln(x)], \]
\[ J_D(x) = J_D[\ln(x)]. \]

(78)

It turns out, however, that in an appropriate four-stimulus task the operation can be identified. For example, Hagerty and Birnbaum (1978) asked subjects to judge (i) "ratios of ratio," (ii) "ratios of differences," (iii) "differences of ratios," and (iv) "differences of differences." They found that the observed judgments for conditions (i), (iii), and (iv) could be explained in terms of a model involving a single scale \( s \), with all comparisons being based on differences of the form

\[ s(a) - s(b) \]

On the other hand, condition (ii) was accounted for by a model based on subjective ratios of differences of scale values:

\[ \frac{s(a) - s(b)}{s(c) - s(d)}. \]

The conclusion is that the scale \( s \) is consistent with the subtraction model of Eq. (77) applied to ratio and difference judgments of stimulus pairs.

Thus, although pairs of stimuli seem to be compared by computing differences, subjects can and do compute ratios, particularly when those ratios involve differences of scale values. This latter observation is consistent with the fact, mentioned earlier, that people are well aware of the distinction between ratios of lengths and differences of those same lengths.

VIII. CONCLUDING REMARKS

Our general knowledge about the conditions under which numerical representations can arise from qualitative data—representational measurement—has grown appreciably during the past 40 years. Such measurement theory has so far found its most elaborate applications in the areas of psychophysics and individual decision making. This chapter attempted both to convey some of our new theoretical understanding and to provide, albeit sketchily, examples of how it has been applied. Of course, much of the detail that is actually needed to work out such applications has been omitted, but it is available in the references we have provided.

The chapter first expounded the very successful probability models for simple binary experiments in which subjects exhibit their ability to detect
and to discriminate signals that are barely detectable or discriminable. These
models and experiments focus on what seem the simplest possible ques-
tions, and yet complexity arises because of two subject-controlled trade-
offs: that between errors of commission and errors of omission and that
between overall error rate and response times. We know a lot about psycho-
metric functions, ROC curves, and speed-accuracy trade-offs, although we
continue to be plagued by trial-by-trial sequential effects that make estimat-
ing probabilities and distributions very problematic. Generalizing the prob-
ability models to more complex situations—for example, general choice,
categorization, and absolute identification—has been a major preoccupation
beginning in the 1980s, and certainly the advent of ample computer power
has made possible rather elaborate calculations. Still, we are always battling
the tendency for the number of free parameters to outstrip the complexity
of the data.

The second major approach, which focused more on structure than
simple order, involved algebraic models that draw in various ways on Höld-
er's theorem. It shows when an order and operation have an additive repre-
sentation, and it was used in several ways to construct numerical representa-
tions. The line of development began historically with empirical operations
that are associative and monotonic, moved on to additive conjoint struc-
tures in which an operation induced on one component captures the trade-
off between components, and most recently has been extended to the work
on homogeneous, finitely unique structures. The latter, which lead to a
wide variety of nonadditive representations, are studied by showing that the
translations (automorphisms with no fixed points) meet the conditions of
Hölder's theorem. In this representation the translations appear as multi-
plication by positive constants. Further generalizations to conjoint struc-
tures with the empirical (not the induced) operations on components and to
structures with singular points make possible the treatment of fairly com-
plex problems in individual decision making. The most extensive applica-
tions of these results so far have been to generalized theories of subjective
expected utility. These new results have not yet been applied in psychophysics
except for relations among groups of translations to study various
matching experiments. Such models lead to homogeneous equations of
degree different from 1.

The apparent similarity of the probability and algebraic models in which
both kinds of representations are invariant under either ratio or interval scale
transformations is misleading. For the probability models in the ordinal
situation this restriction does not reflect in any way the automorphism
group of the underlying structure, which after all is ordinal, but rather
certain arbitrary conventions about the representation of distributions. In
particular, the data are, in principle, transformed so that the error distribu-
tions are Gaussian, in which case only affine transformations retain that
parametric form. This last comment is not meant to denigrate what can be
done with the probability models, which as we have seen is considerable,
especially in binary situations (see section II).

As we have stressed, the field to date has failed to achieve a true melding
of randomness with structure. This failure makes empirical testing difficult
because we usually are interested in moderately structured situations and
invariably our data are somewhat noisy. Exaggerating slightly, we can han-
dle randomness in the ordinal situation—witness sections II and III—and
we know a lot about structure in the ratio and interval scale cases provided
we ignore the fact that the data are always noisy—witness sections IV
through VII, but we cannot treat both together very well.

One result of this bifurcation is notable differences in how we test the
two kinds of models. Those formulating randomness explicitly are ideally
suited to the response inconsistencies that we observe. But because of their
lack of focus on internal structure, they can be evaluated only globally in
terms of overall goodness of fit. The algebraic models suffer from having no
built-in means of accommodating randomness, but they have the advantage
that various individual structural properties—monotonicity, transitivity,
event commutativity, and so on—can be studied in some isolation. This
allows us to focus rather clearly on the failings of a model, leading to
modified theories. One goal of future work must be to meld the two
approaches.

Acknowledgments

This work was supported in part by National Science Foundation grants SBR-9308959 and
SBR-9540107 to the University of California at Irvine. We thank Michael Birnbaum for his
helpful comments and criticisms.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postu-
Dordrecht: Reidel.
Alper, T. M. (1987). A classification of all order-preserving homeomorphism groups of the
Anderson, N. H. (Ed.) (1991a, b, c). Contributions to information integration theory (Vol. 1:


Carroll, J. D., & DeSoto, G. (1990). Fitting a quasi-Poisson case of the GSTUN (General Stochastic Tree UNfolding) model and some extensions. In M. Schader & W. Gaul (Eds.), Knowledge, data and computer-assisted decisions (pp. 93–102). Berlin: Springer-Verlag.


The Representational Measurement Approach to Problems


Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data:
The interrelationship between similarity and spatial density. Psychological Review, 85, 445-463.


Luce, R. D. (1986). Response times. Their role in inferring elementary mental organization. New York: Oxford University Press.


1 The Representational Measurement Approach to Problems


