Psychological Review
1994, Vol. 101, No. 2, 271-277

Thurstone and Sensory Scaling: Then and Now

R. Duncan Luce

Following a brief summary of the major ideas and assumptions of Thurstone's (1927a) "A Law of Comparative Judgment" and a discussion of its historical context, this article reviews the subsequent major developments of this model in the sensory area. They are grouped as (a) response criteria, which Thurstone incorrectly believed to be of little significance, (b) several interpretations of the underlying continuum, (c) extensions of his formalism to choice and absolute identification, (d) the possibility of non-Gaussian distributions, and (e) extensions to multidimensional representations.

Although this famous article of Thurstone's can be read in isolation, it is somewhat misleading to do so. It is really one slice of a complex of closely related ideas covered in some 11 articles, of which 6 appeared during 1927 and, of those, 3 were in the *Psychological Review.* Of these 11 articles, the most general statement of his main idea is found in Thurstone (1927b). However, the easiest way to appreciate the work as a whole, albeit with considerable redundancy, is in the posthumous collection *The Measurement of Values* (Thurstone, 1959), which assembled these 11 plus 16 of his other articles. For those interested in the history of 20th century psychophysics and psychometrics, these articles are certainly essential reading because they lay the groundwork for some 65 years of additional, and on-going, research of significance.

One curiosity in all these articles, which was typical of many but by no means all journal articles of that era, is how little cross referencing there is among his articles or, for that matter, to any other literature. Other people are sometimes mentioned, but only rarely are specific references provided. It is as if all readers were assumed to be familiar with the entire body of relevant publications. By today's standards, Thurstone's articles seem unscholarly.

Synopsis

The article presents two major ideas. The first is the notion that some psychological attributes—those for which judgments of "greater (or more) than" and of "less than" in some intuitive sense—can be thought of as forming a psychological (or subjective) continuum. Moreover, this continuum is, as one would expect, modeled mathematically as a usually infinite interval of real numbers. The second assumption is that any stimulus that elicits the attribute in question—such as a light of some intensity giving rise to a sensation of brightness, or a crime having associated to it a subjective concept of seriousness—does not do so in a rigid fashion, but rather has values that vary from presentation to presentation of exactly the same stimulus under as identical conditions as we know how to contrive. Thurstone spoke of this representation of the stimulus on the psychological continuum as forming a *discriminable process;* today one would say the representation is as a *random variable*[1] (RV).

The intuitions underlying his formal treatment were old, even then. Fechner (1860/1966) had said,

> Even when applied in the same way, one and the same stimulus may be perceived as stronger or weaker by one subject or organ than by another, or by the same subject or organ at one time as stronger or weaker than at another. Conversely, stimuli of different magnitudes may be perceived as equally strong under certain circumstances. Accordingly we ascribe to the subject or organ at one time or another a greater or lesser sensitivity. (p. 38)

Thurstone's insight was not the existence of subjective variability but the fact that it could be modeled explicitly and something useful could be done with the model.

I completed a near-to-final draft of these remarks before rereading my presidential remarks of over 15 years ago to the Psychometric Society at the celebration of the 25th anniversary of the L. L. Thurstone Psychometric Laboratory (Luce, 1977). Although those observations about Thurstone's contributions tended to be more technical than the present remarks, the major thrust is largely unchanged except for some updating.

The following people commented on an earlier draft of this article: F. G. Ashby, J.-C. Falmagne, G. Iverson, L. E. Marks, and J. A. Swets. I appreciate their help.

Correspondence concerning this article should be addressed to R. Duncan Luce, Institute for Mathematical Behavioral Sciences, University of California, Irvine, California 92717.

[1] The random variable (RV) terminology was just becoming common in probability circles at about this time. For example, Kolmogorov's 1933 (German language) monograph on the foundations of probability (Kolmogorov, 1933/1950) introduced the term without ado, suggesting that it was already in use by specialists. I have attempted to track down its origin by contacting authorities both in probability and in the history of probability. Apparently one of the earliest uses was by Cantelli (1916), who defined the concept explicitly in the discrete case, and he used the Italian term *variable casuale,* which, at least now, is translated into English as *random variable.* Several people have expressed the belief that the measure-theoretic definition and the use of the Russian equivalent to *random variable* is attributable to A. Khinchine, but I have not been able to confirm this. It is clear that it was not in common use in the United States in 1927. However, much later, when the term had become relatively common in textbooks, Thurstone still did not use it. Other related equivalences between Thurstone's and modern terminology are: modal discriminable process = mode of the RV; discriminal deviation = difference between a realization of the RV and its mean; discriminal dispersion = standard deviation of RV; discriminal difference = difference between the realizations of two RVs.

We come next to several assumptions that he made and that he viewed as decidedly secondary. Subsequent history has not always sustained him in that judgment.

With considerable caution, he assumed the RVs to be normally (Gaussian) distributed. As an example of his caution,

> The normal probability curve has been so generally abused in psychological and educational measurement that one has reason to be fearful of criticism from the very start in even mentioning it. The only valid justification for bringing in the probability curve in this connection is that its presence can be experimentally tested. The writer has found experimentally that the normal curve was not applicable for certain stimuli. In most of the experiments the distributions are reasonably close to normal. (Thurstone, 1927b, p. 373)

I believe he may have been a bit overly optimistic about our ability in practice to test the normal distribution in the context of his model, as I discuss in some detail later.

With that postulate and assuming that the response is determined by whichever realization of the two RVs is the larger—a less innocent assumption than he believed—he stated[2] the relation between the difference in the means of the RVs, the z score of the probability of selecting the one stimulus as larger than the other, and the variances and correlation of the RVs. This is the law of comparative judgment of his title, Equation 1 of that article. In modern notation one uses μ for the mean, rather than S, which we would use for the stimulus; z, rather than his x, for the z score of the normal unit distribution, and ρ rather than r for the correlation.[3] The law is as follows:

$$\mu_i - \mu_j = z_{ij}(\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)^{1/2}. \tag{1}$$

The balance of his article is devoted to various specializations of the term involving the standard deviations and correlations of the RVs. The reason for introducing these simplifications was, primarily, to reduce the computational problem by having fewer parameters to estimate from a smaller data base. This was necessary because even the simplified computations were near the limits of the (very limited) computing power then available. Indeed, a substantial portion of the literature in the ensuing 30 years focused more on computational tricks—now of limited interest—than on matters of substance.[4]

Impact

The entire work of 11 articles served a major synthesizing function. It provided a simple, easily understood framework within which many empirical phenomena—psychometric functions, thresholds and Weber's law, the scaling of stimuli with no simple underlying physical attribute—all fit nicely together. In many ways, it has continued to be a major component of psychophysical theory, although as readers shall see, there are attempts to understand the source of the RVs in far greater detail than Thurstone contemplated. Moreover, the framework admitted very important extensions not covered in the articles of 1927–1934. This impact on modern psychophysical theory is probably best illustrated by the comprehensive, technical presentations of Falmagne (1985, 1986).

From our current perspective, three features of the model, apparently not explicitly acknowledged by Thurstone in the 1927 era, have played a major role in its subsequent development. First, his framework of RV representations of sensation in no way limits one to analyzing data from binary comparisons or using binary responses, on which all of his articles of this era focus. Eighteen years later Thurstone (1945) recognized that the model could be extended to choices among more than two alternatives, although his formulation of it lacks the elegance of later work (see Bock & Jones, 1968, and references given later).

Second, response criteria played no role; he simply assumed that whichever RV assumed the larger value determined the response. On this he remarked as follows:

> The statement of the law of comparative judgment in the form of equation 1 involves one theoretical assumption which is probably of minor importance. It assumes that all positive discriminal differences (a − b) are judged A > B, and that all negative discriminal differences (a − b) are judged A < B. This is probably not absolutely correct when the discriminal differences of either sign are very small. . . . It is probable that rather refined experimental procedures are necessary to isolate this effect. The effect is ignored in our present analysis. (Thurstone, 1927a, pp. 277–278)

He could not have been more wrong. More on this later.

Finally, all of his publications on this topic were limited to one-dimensional attributes. I say "on this topic" because another of his major areas of research was factor analysis, which, of course, is inherently multidimensional. But when it came to psychophysical analysis, he did not consider multidimensional cases. That, as readers shall see, has become a substantial area of work in recent years. Moreover, it is possible to interpret Fechner as being aware of such generalizations[5]; witness the comment:

> As a start we will consider only sensation, for although the applications of the principle of psychological measurement reach beyond sensation, as will be shown later, sensation provides a starting point under conditions that are the least complicated and most open to direct observation. (Fechner, 1860/1966, p. 46)

In addition to these three generalizations, many psychophysicists and cognitive psychologists have concluded that choice data are simply too lean to allow a very unique specification of the underlying processes without imposing additional demands. The most common of these, which has resulted in considerable elaborations, is that the models should also account for response times as well as choice. This approach is summarized, along with other matters, in Luce (1986).

The remainder of Thurstone's (1927a) article outlines some of the main threads of subsequent developments; my remarks are grouped by the particular limitations of Thurstone's original formulation that have been mentioned, but in the following

---

[2] He did not give the proof, which, however, is elementary.

[3] Greek symbols are now commonly used for the parameters of theoretical distributions.

[4] F. G. Ashby (personal communication, May 1993) suggested that the impact of such specializations may have been more far-reaching than is generally recognized. For example, the computationally simplest case (V for which $\rho_{i,j} = 0$, σ_i independent of i) is the only case for which performance depends monotonically on intermean distance, and this fact may have been the source of the idea of psychological distance and so ultimately of multidimensional scaling.

[5] To some degree, this is sustained by the fact that the second volume of his work deals with higher level cognitive processes (L. E. Marks, personal communication, May 1993).

order: response criteria, the decision continuum, nonbinary procedures, non-Gaussian distributions, and multidimensional representations.

Response Criteria

During World War II and immediately thereafter, engineers working on problems of radar and sonar detection developed a statistical theory of signal detection, one key feature of which was variability in the information underlying decisions about whether or not a target is present (Peterson, Birdsall, & Fox, 1954; van Meter & Middleton, 1954). They viewed target detection as basically a problem in statistical decision making and, as in that literature, one has a trade-off between errors of the two types, I and II or, in the engineering lingo, a trade-off between a hit (the correct identification of a target when it is present) and a false alarm (the assertion of a target when none is present). This trade-off is known in the engineering literature as a *receiver operating characteristic* or ROC. The family of ROCs is a different way of slicing up the space from the statistician's power curves: the trade-offs between discriminability and the probability of failing to detect the signal with the false-alarm rate held fixed at a value. The ROC trade-off is effected by varying the criterion along a single response dimension. (I return later to the engineers' interpretation of it.)

Several psychologists, primarily David M. Green, John A. Swets, and W. P. ("Spike") Tanner, Jr., at the University of Michigan during the early 1950s encountered this theory and recognized that it might be a suitable model for human signal detection, so they imported it—continuing to call it *the theory of signal detectability*—into psychology (Tanner & Swets, 1954) with very long-lasting and profound effects. Among some of the later summary references are Egan (1975), Green and Swets (1966), Macmillan and Creelman (1991), and Swets (1964).

Few experimental psychologists are today unaware of the definition of an ROC, its summary measures d' or the area under the ROC, and the location or criterion parameter β. (For a careful summary of various measures, see Swets, 1986.) More important, in a sense, is the simple recognition that in psychophysical experiments focused solely on matters sensory, it is quite impossible to avoid a decision-making element that is not under the control of the physical stimulus. The literature has shown unambiguously that a variety of motivational factors—instructions, payoffs, and probability distribution of signal presentations—affect where the subject ends up on the ROC determined by the stimuli. The subject establishes a response criterion whose value can be influenced by nonsensory experimental variables.

This important development took place apparently quite oblivious of Thurstone's earlier work,[6] although the underlying RV model was substantially the same as his. The connection was, of course, quickly pointed out by others once it entered the psychological literature, and it was a significant advance over Thurstone's work. It said, contrary to his belief, that the response criterion is a major feature, not an insignificant problem, of psychophysics. It was about this belief that I earlier commented, "[Thurstone] could not have been more wrong."

The Decision Continuum

One obvious question is: Just what is the continuum on which the representation occurs? Is it merely a formal (hypothetical) construct of theorists or does it have some more substantial existence? There is no firm consensus on this, even today. Some hold it to be simply a hypothetical construct with no more or less substance than the behavioral data themselves, and so in this view the continuum and RVs simply provide a reasonably compact summary of the data.

Others hope—it is really no more than that at present—that the continuum corresponds to something actually going on in the nervous system. Attempts exist to correlate gross measures of electrical activity of the brain with detection behavior and to conjecture that the electrical activity is an index of the continuum (R. Galambos, 1974; Squires, Hillyard, & Lindsay, 1973). Others have attempted somewhat detailed, but hypothetical neural models in which trains of "neural pulses," usually modeled as Poisson or other closely related renewal processes, are processed to arrive at some number on which the response is based. In many cases, this number is an estimate of the intensity parameter of the Poisson process. For example, counting and timing models have been described in which it is assumed that the brain estimates signal intensity as either the number of counts to occur in a fixed time period or the amount of time it takes to get a fixed count (Luce & Green, 1972; McGill, 1967; McGill & Teich, 1990).

Such models not only account for the choice behavior but provide an account of the time it takes to respond in different situations. Green and Luce (1973) argued that in natural environments, where there is enormous uncertainty about the intensities of significant signals, it is plausible that organisms use timing. However, in most experimental designs, signals of fixed, known duration are used, and counting is appreciably more effective than timing. The training needed to achieve asymptotic choice behavior may arise from the reprogramming needed to switch from timing to counting. This led to two different experimental procedures, one to invoke counting and the other timing, which in turn led to sharply differential predictions about both the ROC curve and the speed–accuracy trade-off function for the two cases. Their auditory data provided supporting evidence that people can be induced to exhibit either behavior (Luce, 1977, provides a summary). Wandell (1977) established similar empirical results for visual detection.

The third interpretation was that of the engineers who developed the theory of signal detectability. In their situation, the signals along with their accompanying masking noise had very complex physical representations that in no way could be viewed as one-dimensional. Nevertheless, the decision process must somehow convert this mass of information into a form suited to a binary decision. They suggested that it be recast and simplified as the statistical likelihood ratio: the ratio of the

---

[6] I say "apparently" because John A. Swets (personal communication, April 1993) was, in fact, quite aware of Thurstone—Clyde Coombs, a professor at Michigan and Thurstone's student, made sure of that. However, because the RV representation idea was by then fairly widespread (see Swets, 1973, for a fairly detailed history), they felt it unnecessary to mention Thurstone explicitly.

probability density of there being a target to the density of there not being a target, both densities being conditional on the observed data. They then argued, invoking the central limit theorem, that the distributions of likelihood ratio under the two hypotheses, target (signal) and no target (noise alone), should be Gaussian.

Many psychologists find this account of the continuum and Gaussian distributions psychologically quite implausible because of the amount of prior data that must be accumulated to build up the distributions that are involved. Such data could be collected empirically in at least some engineering settings, but it is difficult to imagine how a functioning organism operating in a complex and variable environment is accumulating this vast store of highly specific information. By and large, psychologists have opted for one or the other of the first two accounts.

Nonbinary Procedures

Choice

As was noted earlier, Thurstone (1945) recognized that the original binary model, which really describes the choice of one of two alternatives, can be extended to choices among more than two alternatives. My impression is that this article was little noticed, and it was not until somewhat later that economists, then unaware of Thurstone, became interested in what they (and the subsequent literature) called the *random utility model,*[7] although it is in no way limited to the case of utility, that is, the attribute of preference. Denote by S a set of stimulus alternatives that are thought to vary in one subjective attribute, and suppose that for s in S there exists an RV representation $\mathbf{X}_s$. Then the probability $P(s;S)$ that s will be selected from S as possessing the largest degree of the attribute in question is described simply by the probability that its random variable assumes the largest value, that is,

$$P(s;S) = P(\mathbf{X}_s = \max_{r \in S}\{\mathbf{X}_r\}). \tag{2}$$

Within economics, the RVs are usually written $\mathbf{U}_s$ or U(s) to emphasize the utility interpretation. The simplest version of this random utility model, and hence the one most studied, assumes independent RVs. I suspect that this was the model Thurstone was describing in his 1945 article, although I find it somewhat difficult to tell for sure. In any event, a large literature has ensued, which I make no attempt to survey here. For more details, see Suppes, Krantz, Luce, and Tversky (1989, chap. 17).

Absolute Identification

As well as one can tell from his published articles, Thurstone did not see the fairly natural generalization of his model to procedures of absolutely identifying (AI) one of n stimuli that vary on a single attribute. In several articles he was quite explicit that the several stimuli should be viewed as represented by RVs on the same continuum. What Thurstone apparently did not see, or care about, was the possibility that in AI situations involving n signals, subjects may establish $n - 1$ criteria designed to partition the subjective continuum into intervals corresponding to the several stimuli. The presentation on a trial is then identified as stimulus s_i if the observed value of the RV falls in the ith interval. Apparently, the first person to have stated this model explicitly was Torgerson (1954) in what, for most psychologists, was a somewhat obscure outlet; his more widely read book (Torgerson, 1958) provided a full statement of the model for AI. He arrived at equations, analogous to Equation 1, which he called the *law of categorical judgment.* This added, of course, a minimum of $n - 1$ additional parameters, double that if the criteria are assumed to be variable, which seems fairly plausible, and still more if they are correlated with each other or with the stimuli. Thus, with computing power still being very limited, much attention was devoted to easing the computational burden. I need not go into that here.

It is clear that if the standard deviations of the stimulus RVs are all the same and those of the criterion RVs are also all the same, then there is no way from simple AI data to identify how the variability partitions between the two sources; the one can be increased at the expense of decreasing the other without affecting the observed probabilities. Even when the variances vary, separating them computationally is fairly tricky (Braida & Durlach, 1972; Durlach & Braida, 1969). Thus, some effort has gone into devising ways to see the partitioning. For example, Nosofsky (1983) pointed out that if on each trial a signal is repeated N times, then the performance measure $d'(N)$ and N are related by the formula

$$[1/d'(N)]^2 = \sigma_c^2 + \sigma_s^2/N, \tag{3}$$

where σ_c^2 is the criterion variance and σ_s^2 is the signal variance. An experiment confirmed this linear relation and showed both variances to be substantial. Moreover, increasing the range of other signals increased the values of the estimates of both the signal and criterion variances.

A possibly related problem was first given prominence by Miller (1956) who, however, cast it somewhat indirectly in terms of the, then fashionable, information-theoretic measure of uncertainty or entropy. The result is simply described. Consider the peripheral representation $\mathbf{X}_s$ of a particular signal s, such as a light patch or a pure tone. It seems plausible that under well-controlled experimental conditions and with sufficient time between trials, such peripheral representations are totally independent of the experimental demands made on the subject. From what researchers know about the peripheral nervous system, activity there depends on the stimulating conditions but not on the question to which the subject is attempting to respond or to the pattern of stimuli on other trials.[8] In particular,

---

[7] J.-C. Falmagne (personal communication, April 1993) emphasized the different interpretation of the underlying sample spaces for psychologists and economists, with the former interpreting the RVs as a property of individuals with the space as some sort of internal continuum, and the latter thinking of the RVs as projections over random utility vectors with the space being the individuals. In practice, the distinction is not always maintained in the sense that psychologists often estimate probabilities over subjects.

[8] It is surprising that there do not appear to be any directly relevant data on this point, Wasserman (1991) notwithstanding. He interpreted data of Knibestöl and Vallbo (1980) as showing a range effect on the mean exponent of power functions fit to firing rate versus intensity data. However, these data were obtained from different units and so one cannot tell whether the effect is due to the stimulus range or to the (small, 23, 16, and 21) samples of afferent units which can be clearly seen, from

$\mathbf{X}_s$ should be the same whether the subject is trying to identify either 1 of 2 possible signals or 1 of 10 possible signals.

The behavioral data are otherwise. For example, separating two pure tones by 5 dB is more than enough to result in perfect identification, whereas to make 10 signals, equally spaced in dB, perfectly identifiable requires separating adjacent ones by about 15 dB. Such degeneration of performance as the number and range of the signals increases is true for intensity in all sensory modalities. Of great conceptual interest is the question of whether the signal variance or the criterion variance, or both, increase in going from 2 to 10 stimuli. If the former variance depends on the number of stimuli and if there is no evidence of any change in the peripheral information available, then one can only conclude that there is some sort of central filtering that, on average, reduces the amount of information, and so increases the variance, on which the decision is based. If the criterion variance grows, then some attempt must be made to understand why this is coming about.

One proposal was an attention band model of Luce, Green, and Weber (1976) that postulates that the central system can only monitor fully limited subsets of peripheral neurons, the rest being far less fully monitored (see also Luce & Green, 1978). Another line of thought is that the criteria are adjustable on the basis of local, trial-by-trial experience, but when data are averaged over trials, as they typically are, the adjustments appear as an apparent increase in variability. Limited attempts have been made to model such adjustment processes, although in practice it has proved quite difficult to test them in any very decisive way. Some of the relevant references are Marley and Cook (1984, 1986), Treisman (1985), and Treisman and Williams (1984).

Of course, the Nosofsky result described earlier suggests that both mechanisms are at play, which only complicates the problem. This fact, along with the possibility of non-Gaussian distributions, seems to be ignored by some critics of the idea of an attention band (Kornbrot, 1980).

Non-Gaussian Distributions

Thurstone clearly recognized that in assuming Gaussian distributions he was making an inessential, if convenient, assumption. The question, of course, was what other distributions to assume and why. Depending on one's exact interpretation of the continuum, various possibilities can arise. I cite two quite different examples with which I was involved.

Within the detection context, Wandell and Luce (1978) observed[9] that averaging neural counts was not the only calculation that the nervous system might carry out to arrive at decisions. Empirical estimates of observed neural firing rates and of behavioral response times and accuracy had made it unambiguously clear that the counts must accumulate over many, perhaps several hundred, parallel sources of information. Thus, instead of averaging these counts, the system could just as well arrive at a decision by observing the slowest count. Just as the central limit theorem states that under very general conditions the average of similar, independent RVs converges to a unique distribution—the Gaussian—an analogous but, to psychologists, less well-known result (see J. Galambos, 1978, or Luce, 1986, for expositions) says that the distribution of the maximum of independent, identically distributed RVs converges to one of three types of distribution, which one depending on the nature of the upper tail of the underlying distribution. In the Poisson case, which we assumed, that asymptotic distribution is the double exponential:

$$P(\mathbf{X}_s \leq x) = \exp[-\exp(-\alpha_s(x - \beta_s))]. \tag{4}$$

Although this distribution is unimodal, it is appreciably different from the Gaussian. Nonetheless, as Wandell and I showed, ROC data could be fit equally well by either model. Contrary to Thurstone's belief, expressed in the quotation on p. 272, it was not possible to decide easily between these two distributions just on the basis of the accuracy of the responses. A major difference between the models was predicted for the speed–accuracy trade-off function, and the data we collected unambiguously favored the Gaussian over the double exponential distribution.

A second way in which alternative distributions have arisen is in considering the connection between the random utility model for choices and the choice model (which generalized Bradley & Terry's, 1952, binary model) in which there is a function v over the alternatives such that

$$P(\mathrm{s};\mathrm{S}) = v(\mathrm{s})/\Sigma_{\mathrm{r}\in\mathrm{S}}v(\mathrm{r}). \tag{5}$$

Luce (1959) provided a purely behavioral axiom in terms of the choice probabilities that is easily shown to be equivalent to this representation. For the binary case, Bradley and Terry (1952) showed it was related to the logistic distribution; for the general choice model, E. Holman and A. A. J. Marley (as reported in Luce & Suppes, 1965) first noted that it is equivalent to an independent random utility (Thurstonian) model, Equation 2, in which the RVs are exponentially distributed. There is, of course, nothing terribly unique about this family of distributions, except for the important fact that they cannot be transformed into Gaussian distributions. The reason for the nonuniqueness is that if, for s in S, $\mathbf{X}_s$ are RVs and f is any strictly increasing function, then $\mathbf{Y}_s = f(\mathbf{X}_s)$ are equally well RVs with the property that the maximum remains invariant. For example, letting $\mathbf{Y}_s = -\log \mathbf{X}_s$, one can show that if $\mathbf{X}_s$ is distributed exponentially, then $\mathbf{Y}_s$ is distributed as the double exponential with $\alpha = 1$ and $\beta(\mathrm{s}) = v(\mathrm{s})$. (See Suppes et al., 1989, p. 425.)

A natural question to be raised is whether some inherently different distribution, not transformable into the double exponential, is compatible with the choice model. Laha (1964) showed this to be possible, but McFadden (1974) and, independently, Yellott (1977) showed that if the presence of repeated copies of alternatives in the choice set did not disrupt the choice pattern, then the double exponential is, except for monotonic transformations, unique.

Additional results about random utility models are described by McFadden (1981) and Suppes et al. (1989).

---

their Figure 2, to have highly variable exponents. One needs data from the same units that are run under different conditions.

[9] In a somewhat more general context, Thompson and Singh (1967) showed the same thing, but they did not pursue it experimentally.

Multidimensional Representations

As I observed earlier, Thurstone did not give serious consideration in the context of these psychophysical models to stimuli that vary on two or more sensory attributes, and that the engineers assumed that they could effectively reduce their clearly multidimensional signals to the one-dimensional decision criterion of likelihood ratio. Although some minor attempts were made within the context of the theory of signal detectability to treat signal representations as multidimensional (Tanner, 1956), it was difficult to provide a sensible interpretation of two or more axes in terms of likelihood ratios. The generalization of the theory of signal detectability to the absolute identification of several signals—one of which is often no signal—has not been very effective.

If, however, one abandons the idea that the continuum is likelihood ratio and treats it as some sort of sensory attribute, then it is not difficult to think of each stimulus as being described by a multivariate distribution. For example, suppose the stimuli are pure tones that vary in both frequency and intensity, then it is plausible to think of the stimuli as being represented in a subjective space of, for example, pitch and loudness, where the representation is usually assumed to be a bivariate, possibly correlated, Gaussian distribution. Various tasks are possible. Absolute identification is the most refined identification task; various more-or-less elaborate forms of categorization are less refined ones. It is clear that solving any of these tasks within this framework entails some partitioning of the subjective space. Moreover, here is the rub: If curves of any degree of complexity are permitted, the number of parameters tends to get totally out of control. Thus, there is a tension between increasing the number and complexity of tasks and in restricting the freedom on how the space is partitioned to maintain sufficient degrees of freedom to test the model. In recent years there have been increasing efforts in this realm; some sense of the approaches taken can be found in the recent edited volume of Ashby (1992). In particular, identifiability (in principle) has been established if the decision bounds are assumed to be quadratic (Ashby, 1992, p. 462).

Conclusions

Although "A Law of Comparative Judgment" is justly famous, it is just one piece of the body of work called Thurstonian scaling, which has dominated thought about sensory scaling since then. Within that framework, psychologists have constructed many new developments that extend its reach to many areas. To my mind, the most important of these have been (a) the broadening of the theory from binary choices to general choices and to general absolute identification and categorization procedures, including random vector representations; (b) the great emphasis since 1950 on response criteria and, in particular, on the ROC; (c) the increasingly detailed studies about the source of the sensory RVs that, among other things, has led to non-Gaussian distributions; and (d) the related development of models that attempt to account for response time as well as choice.

Thurstone stands as one of the great synthesizers of sensory scaling in the 20th century, and his synthesis remains alive and is increasingly being refined.

References

Ashby, F. G. (Ed.). (1992). *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice.* San Francisco: Holden-Day.

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

Braida, L. D., & Durlach, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483–502.

Cantelli, F. (1916). Sulla legge dei grandi numeri [On the law of large numbers]. *Atti della R. Accademia dei Lincei. Memorie della Classe di Scienze Fisiche, Matematiche, e Naturali, ser. 5, 9,* 330–349.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372–383.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* San Diego, CA: Academic Press.

Falmagne, J.-C. (1985). *Elements of psychophysical theory.* New York: Oxford University Press.

Falmagne, J.-C. (1986). Psychophysical measurement and theory. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 1-1–1-66). New York: Wiley.

Fechner, G. (1966). *Elements of psychophysics, Vol. I* (H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

Galambos, J. (1978). *The asymptotic theory of extreme order statistics.* New York: Wiley.

Galambos, R. (1974). The human auditory evoked potential. In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement* (pp. 215–221). Dordrecht, Holland: Reidel.

Green, D. M., & Luce, R. D. (1973). Speed-accuracy trade-off in auditory detection. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4, pp. 547–569). San Diego, CA: Academic Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Knibestöl, M., & Vallbo, Å. B. (1980). Intensity of sensation related to activity of slowly adapting mechanoreceptive units in the human hand. *Journal of Physiology, 300,* 251–267.

Kolmogorov, A. (1950). *Foundations of the theory of probability* (N. Morrison, Trans.). New York: Chelsea. (Original work published 1933)

Kornbrot, D. (1980). Attention bands: Some implications for categorical judgment. *British Journal of Mathematical and Statistical Psychology, 32,* 1–16.

Laha, R. G. (1964). On a problem connected with beta and gamma distributions. *Transactions of the American Mathematical Society, 113,* 287–298.

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

Luce, R. D. (1977). Thurstone's discriminal processes fifty years later. *Psychometrika, 42,* 461–489.

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

Luce, R. D., & Green, D. M. (1972). A neural timing theory for response times and the psychophysics of intensity. *Psychological Review, 79,* 14–57.

Luce, R. D., & Green, D. M. (1978). Two tests of a neural attention hypothesis in auditory psychophysics. *Perception & Psychophysics, 23,* 363–371.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention band in absolute identification. *Perception & Psychophysics, 20,* 49–54.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Hand-*

book of mathematical psychology (Vol. 3, pp. 249–410). New York: Wiley.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology, 37,* 136–151.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychophysical judgments applied to magnitude estimation. *Journal of Mathematical Psychology, 30,* 339–390.

McFadden, D. (1974). Conditional logit analyses of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). San Diego, CA: Academic Press.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data* (pp. 198–272). Cambridge, MA: MIT Press.

McGill, W. J. (1967). Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology, 4,* 351–376.

McGill, W. J., & Teich, M. (1990). Auditory signal detection and amplification in a neural transmission network. In M. Commons & J. Nevin (Eds.), *Signal detection* (pp. 1–37). Hillsdale, NJ: Erlbaum.

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

Nosofsky, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute judgment. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 299–309.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *IRE Transactions, Professional Group on Information Theory, 4,* 171–212.

Squires, K. C., Hillyard, S. A., & Lindsay, P. H. (1973). Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. *Perception & Psychophysics, 13,* 25–31.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement, Vol. 3.* San Diego, CA: Academic Press.

Swets, J. A. (Ed.). (1964). *Signal detection and recognition by human observers.* New York: Wiley.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182,* 990–1000.

Swets, J. A. (1986). Indices of discrimination and diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

Tanner, W. P., Jr. (1956). Theory of recognition. *Journal of the Acoustical Society of America, 28,* 882–888.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision making theory of visual detection. *Psychological Review, 61,* 401–409.

Thompson, W. A., Jr., & Singh, J. (1967). The use of limit theorems in paired comparison model building. *Psychometrika, 32,* 255–264.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review, 34,* 273–287.

Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology, 38,* 368–389.

Thurstone, L. L. (1945). The prediction of choice. *Psychometrika, 10,* 237–253.

Thurstone, L. L. (1959). *The measurement of values.* Chicago: University of Chicago Press.

Torgerson, W. S. (1954). A law of categorical judgment. In L. H. Clark (Ed.), *Consumer behavior* (pp. 92–93). New York: New York University Press.

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

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology, 29,* 175–230.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

van Meter, D., & Middleton, D. (1954). Modern statistical approaches to reception in communication theory. *IRE Transactions, Professional Group on Information Theory, 4,* 119–145.

Wandell, B. (1977). Speed–accuracy tradeoff in visual detection: Applications of neural counting and timing. *Vision Research, 17,* 217–225.

Wandell, B., & Luce, R. D. (1978). Pooling peripheral information: Average vs. extreme values. *Journal of Mathematical Psychology, 17,* 220–235.

Wasserman, G. S. (1991). Neural and behavioral assessments of sensory quantity. *Behavioral and Brain Sciences, 14,* 192–193.

Yellott, J. I., Jr. (1977). The relationship between Luce's axiom, Thurstone's theory of comparative judgment and the double exponential distribution. *Journal of Mathematical Psychology, 15,* 109–144.

Received July 6, 1993
Accepted September 23, 1993 ■