

Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment

DIANA EUGENIE KORNBROT
The Hatfield Polytechnic, Hatfield, Herts., England

Theoretical predictions of the two models are analyzed and shown to be different if the number of categories is three or more. Then two common methods of testing mathematical models empirically are examined: (1) direct testing of predictions embodied in constraint equations, and (2) minimum chi-square methods. In the experimental investigation, two undergraduate subjects performed an eight-stimulus, 8-response category judgment of loudness in a neutral and a payoff biased condition. There were no significant departures from the Thurstone model, but substantial and significant departure from the choice model, especially in the biased condition. Chi-square tests were more powerful than constraint tests, but less good at identifying the source of conflict between data and theory.

The categorization of perceived objects on a single dimension has been of interest to psychologists since the pioneering work of L. L. Thurstone in the 1920s. The appeal of the problem lies in the fact that it is rich enough to capture interesting features of mental processing and, at the same time, systematic enough to be amenable to theoretical analysis. The two main such analyses are the original formulation of Thurstone (1927) and the choice model of Luce (1959, 1963, as applied to categorical judgment). The first aim of the present paper is to find the best possible model of categorical judgment. It will be shown that the two models make different theoretical predictions, and that the predictions of the Thurstone model are confirmed by experiment, whereas the predictions of the choice model are significantly in error. The second aim is a methodological one of assessing available techniques for testing alternative mathematical models in psychology, using choice and Thurstone models of category judgment as a concrete example.

In a general category judgment experiment, there are n different types of stimuli, S_i , and m different responses, R_g . For example, in a psychological task such as the one used to test the model here, the stimuli are pure tones with n different loudnesses and the responses are the digits 1 to m . However, a structurally identical experiment could be constructed with, say, people as the stimuli and psychological

health on a scale from 1 to m as the response. The experimenter presents each type of stimulus to a subject many times in random order. On each presentation, the subject makes a response. Then the empirical result of the experiment is a judgment matrix $\{o_{ig}\}$, where the o_{ig} are the observed frequencies of response g to stimulus i . The experimental task is termed absolute identification (AI) when the number of stimuli, n , is equal to the number of responses, m , as in the present study. It is termed a rating experiment when $n < m$ and a general category experiment when $n > m$. The Thurstone model is formulated in such a way as to apply whatever the relation between m and n . The choice model can be reformulated so as to apply for all values of n, m (Jacobs, 1974). However, the techniques are cumbersome and not worth repeating, since our results show that the choice model is not applicable to even the AI form of categorical judgment.

Both models assume that the $n(m-1)$ degrees of freedom in the empirical judgment matrix may be accounted for by only $n-1$ sensory parameters and $m-1$ motivational parameters. The $n-1$ sensory parameters correspond to the ease of discrimination of the $n-1$ adjacent stimulus pairs, while the $m-1$ motivational parameters correspond to the $m-1$ partitions the subject must make to assign the stimulus objects to m different categories. Since there are more degrees of freedom (df) in the empirical judgment matrix than there are theoretical parameters, a number of independent constraints, I , between cells of the judgment matrix must exist. I is given by the df in the empirical matrix minus the number of theoretical parameters. The theoretical section covers the derivation of the theoretical parameters and the construction of constraint equations from the judg-

Requests for reprints should be sent to D. E. Kornbrot, The Psychology Group, The Hatfield Polytechnic, P. O. Box 109, Hatfield, Herts., England. Grateful thanks are extended to Sean McKeown and other members of The Hatfield Polytechnic Numerical Optimisation Centre for invaluable help in using the optimization routines and fruitful discussions on general methods of optimization.

ment matrix for both theoretical models. It is then shown that, for $n > 2$, the constraints for the choice model conflict with the constraints for the Thurstone model.

It is not sufficient to show that the theoretical predictions of the two models are different. One has first to design an experiment where the predicted differences will be large enough to be recognized.

Then one has to develop some form of analysis which will determine whether the differences between theory and experiment could have arisen from random errors of measurement or whether they are, indeed, "significant." The approach of mathematical modellers has been necessarily ad hoc since the test applied is usually an integral part of the model, a situation which makes it very difficult to compare models that are differently formulated. For example, it is quite unlikely that the power of the tests of the two models will be the same, even when both models are tested at the same significance level. In the section on empirical comparison of models, we take two methods which are commonly used for testing specific mathematical models and try to abstract the general features which would be common to any model tested. There is probably not much in this discussion which was discovered less than 100 years ago. However, this material tends not to be generally available to psychologists in the form of a description of methodology applicable to any problem in the way that, for example, the analysis of variance is described in textbooks of statistics for psychologists.

The first method investigated is the method of constraint equations. Frequently, the axioms of a model are embodied in a simple equation of the form $T(x_0)$ is a constant, where the x_0 are quantities measured in an experiment. Tests of the axioms of decision theory are frequently of this type (see Edwards & Tversky, 1967, for articles of this kind). To determine whether T is, indeed, significantly different from the prescribed constant, one must have some method for estimating the errors of measurement of T , given, for example, some estimate of the errors in the individual observed parameters, x_0 . The theory of propagation of errors, familiar to physical scientists (Martin, 1971), is expounded and applied to the problem of testing the constraints derived in the theory section.

The second method discussed is the optimization method for determining the best values of the theoretical parameters of a model and then testing whether the observed difference between theory and experiment is statistically significant. This approach has often been used to test Thurstone models, starting with Gullicksen (1954) and progressing through to, for example, the sophisticated program developed by Arbuckle and Nugent (1973). However, none of the available programs could have been applied to

the choice model without very considerable modification. For this reason, a general optimization routine was used. It is interesting to note that no such general-purpose optimization routines exist within SPSS or the BIOMED package of computer routines. As with the constraint methods, the optimization methods are widely used to test specific models, but it is quite hard to discover a general discussion of the methodology involved.

The discussion of empirical comparison methods concludes with an assessment of the relative merits of the two approaches.

THEORETICAL MODELS

The Choice Model for n Stimuli and m Responses

The basic intuitions underlying the choice model are described by Luce (1959, 1963, p. 113). As with other modern theories of categorization, there are two kinds of parameter: sensory and motivational. Both sensory and motivational influences are described in terms of independent probabilities which quantify two distinct tendencies of decision makers. The tendency to confuse similar stimuli is described by the parameter u_{ig} , which is the probability of response g to stimulus i in the absence of response bias. The u_{ig} are the sensory parameters of the choice model, and they provide a sensory scale which obeys the choice axiom. Motivational tendencies are described by response bias parameters, b_g , which give the probability of response g in the absence of any knowledge of which stimulus was presented. A crucial assumption of the model is that the u values are independent of the b values. Then, using the laws of combination of independent probabilities, the probability of response g to stimulus i , p_{ig} , is given by:

$$p_{ig} = k_i b_g u_{ig} = k_i a_{ig}. \quad (1)$$

Then empirical frequency, o_{ig} , is given by $N_i \cdot p_{ig}$, where N_i is number of presentations of stimulus i . The matrix $\{a_{ig}\}$ is called the scale value matrix (Luce, 1963). The k_i are constants of proportionality which ensure that each row of the matrix sums to unity, as the observer makes some response to each stimulus presentation. Hence,

$$k_i = \sum_{g=1}^{g=m} b_g u_{ig}. \quad (2)$$

The interpretation of the u as a scale of similarity rests on the u_{ig} obeying the three choice axioms as described by Luce (1963).

$$\text{Axiom 1} \quad u_{ig} = u_{gi} \text{ for all } i, g.$$

Similarity is an essentially symmetric concept.

Axiom 2 $u_{ij} = 1.$

The importance of this axiom lies in the fact that the similarity of any stimulus to itself is the same as the similarity of any other stimulus to itself. This constant similarity is arbitrarily put equal to unity for normalization purposes.

Axiom 3 $u_{ik} \geq u_{ij} \cdot u_{jk}$ for all $i, j, k.$

With these three axioms, the similarity scale is a ratio scale. Furthermore, if a derived scale, called the *d*-scale, is defined by:

$$d_{ig} = -\log(u_{ig}), \tag{3}$$

then the *d*-scale is an interval scale which obeys the following three distance axioms which are equivalent to the above three choice axioms.

Axiom 1a $d_{ig} = d_{gi},$

Axiom 2a $d_{ii} = 0,$

Axiom 3a $d_{ik} \leq d_{ij} + d_{jk}.$

This distance form of Axiom 3 is the familiar triangle inequality. The equality holds if the *d*-scale is unidimensional and stimulus *j* lies between stimuli *i* and *k*. (In this event, the equality of course holds in Axiom 3 as well as Axiom 3a.)

The parameters of the choice model may then be calculated from the scale value matrix $\{a_{ig}\}$:

$$a_{ig} = b_g \cdot u_{ig}. \tag{5}$$

So the scale value matrix, a_{ig} , has the form:

Stimulus	Response					
	"1"	"2"	...	"g"	...	"m"
s_1	$b_1 u_{11}$	$b_2 u_{12}$...	$b_g u_{1g}$...	$b_m u_{1m}$
s_2	$b_1 u_{21}$	$b_2 u_{22}$...	$b_g u_{2g}$...	$b_m u_{2m}$
...						
s_i	$b_1 u_{i1}$	$b_2 u_{i2}$...	$b_g u_{ig}$...	$b_m u_{im}$
...						
s_n	$b_1 u_{n1}$	$b_2 u_{n2}$...	$b_g u_{ng}$...	$b_m u_{nm}$

(6)

This formulation is suggested by Luce (1963, p. 175, for recognition) and by Luce and Galanter (1963, p. 271, for category scaling). Luce also imposes the constraint that $n = m$ and then suggests that estimates

of the $u_{i,i+1}$ and the ratios b_{g+1}/b_g may be calculated by picking 2×2 submatrices from the full probability matrix $\{p_{ig}\}$, using only diagonal and adjacent entries.

If one extends Luce's approach to utilize elements more than one step away from the diagonal, then all the u_{ij} and all the ratios b_{g+h}/b_g may be calculated. Sets of 2×2 submatrices are selected from diagonal elements and elements symmetrically placed *h* units from the diagonal. Such scale submatrices have the form:

$$\begin{matrix}
 s_g & & g & & g+h \\
 & & \left[\begin{array}{cc} b_g u_{gg} & b_{g+h} u_{g,g+h} \\ b_g u_{g+h,g} & b_{g+h} u_{g+h,g+h} \end{array} \right] & & \\
 s_{g+h} & & & &
 \end{matrix} \tag{7}$$

and are denoted here *h*-tuple separation matrices. For a particular value of *h*, there are $(m - h)$ *h*-tuple separation matrices, enabling one to calculate $(m - h)$ sensory parameters, $u_{g,g+h}$, and $(m - h)$ ratios of bias parameters, b_{g+h}/b_g .

From Axiom 1

$$u_{gg} = u_{g+h,g+h} = 1, \tag{8}$$

and from Axiom 2

$$u_{g,g+h} = u_{g+h,g}. \tag{9}$$

Then, equating ratios of terms from Theoretical Submatrices 7 with corresponding ratios from the empirical judgment matrix, one obtains:

$$u_{g,g+h}^2 = \frac{p_{g,g+h} p_{g+h,g}}{p_{gg} p_{g+h,g+h}} \tag{10}$$

and

$$\left[\frac{b_{g+h}}{b_g} \right]^2 = \frac{p_{g+h,g} p_{g+h,g+h}}{p_{gg} p_{g,g+h}}. \tag{11}$$

Equation 11 has previously been proposed by Shepard (1957, Equation 38). It determines the motivational parameters, b_g , up to a multiplicative constant. The constant is set by the normalization convention, either

$$\sum_{g=1}^{g=m} b_g = 1 \tag{12}$$

or by putting $b_1 = 1$, as in Luce's treatment of the 2×2 case and the present study. Equations 10, 11, and 12 give a full solution for the $n - 1$ independent bias parameters and $\frac{1}{2}n(n - 1)$ independent stimulus parameters.

Since the empirical probability matrix has $n(n - 1)$ df, while the choice model has only $\frac{1}{2}(n - 1)(n - 2)$ independent parameters, there must be a number, I , of constraint equations relating the elements of the empirical judgment matrix. These constraint equations may be obtained by noting that the ratio of b_{g+h}/b_g is given by a product of ratios of b_g s from adjacent responses as expressed in Equation 13:

$$\frac{b_{g+h}}{b_g} = \frac{b_{g+h}}{b_{g+h-1}} \frac{b_{g+h-1}}{b_{g+h-2}} \dots \frac{b_{g+1}}{b_g} \quad (13)$$

Ratios of adjacent bias parameters for substitution in Equation 13 are obtained from the single separation ($h = 1$) version of Equation 11. Then, equating estimates of the b_{g+h}/b_g from the h separation version of Equation 11 and from Equation 13, one obtains the $\frac{1}{2}(n - 1)(n - 2)$ independent constraint Equations 14:

$$\frac{p_{g,g+h} p_{g+h,g+h-1}}{p_{g,g+1} p_{g+h,g}} \prod_{i=g+2}^{i=g+h} \frac{p_{i-1,i-2}}{p_{i-1,i}} = 1. \quad (14)$$

Equation 14 is equivalent to the product rule for binary preference of Luce and Suppes (1965, Equation 27). It is expressed here in the form of products of ratios of probabilities from the same row of the judgment matrix. This is not the simplest form, but it is the most convenient for estimating errors, as elements which come from the same row, and are thus dependent on each other, are grouped together. That each of the Equations 14 is independent of all the others may be confirmed by starting with $h = 1$ and working from $g = 1$ to $g = n - 1$ and observing that each equation has at least one element that has both not been used before and is not determined by the normalization convention. One may then repeat the process for the $h = 2$ set, with values of g from 1 to $n - 2$, and so on. The maximum possible number of independent constraints, I , is given by the df in the empirical matrix minus the number of estimated parameters; thus:

$$\begin{aligned} I &= n(n - 1) - [(n - 1) + \frac{1}{2}n(n - 1)] \\ &= \frac{1}{2}(n - 1)(n - 2). \end{aligned} \quad (15)$$

So the constraints implied by Equations 14 for $h \geq 2$ must exhaust the maximum possible number of constraints.

In actual experiments, most of the data fall on or near the diagonal of the judgment matrix, so the model will actually be tested only with $h = 2$; then Equation 14 becomes:

$$\begin{aligned} (p_{g,g+2}/p_{g,g+1})(p_{g+2,g+1}/p_{g+2,g})(p_{g+1,g}/p_{g+1,g+2}) \\ = 1 \end{aligned} \quad (16)$$

or, in additive form:

$$\begin{aligned} B_g &= \ln p_{g,g+2} - \ln p_{g,g+1} + \ln p_{g+2,g+1} \\ &\quad - \ln p_{g+2,g} + \ln p_{g+1,g} - \ln p_{g+1,g+2} = 0. \end{aligned} \quad (17)$$

The B_g are denoted choice-bias parameters. If they are zero, then the bias parameters of the choice model are internally consistent.

It is interesting that Equations 10, which determine the sensory parameters, Equations 11 and 12, which determine the bias parameters, and Equations 14, which determine the constraint equations, have all been obtained from the general formulation without using Axiom 3. The crucial postulates in these derivations are Axioms 1 and 2 and the assumption that bias and sensory parameters are independent and can be combined in the same way as uncorrelated probabilities (cf. Equation 1).

The constraint equation method may also be used to test the validity of Axiom 3. Luce (1963) suggests that if the stimuli are unidimensional and the equality form of Axiom 3 holds, then the axiom may be used recursively to estimate the separation, u_{ij} , between stimulus i and stimulus j from the separation between adjacent stimuli. Thus,

$$u_{ij} = u_{i,i+1} u_{i+1,i+2} \dots u_{j-1,j}. \quad (18)$$

The u_{ij} s for adjacent stimuli are obtained from the single separation version of Equation 10. Equations 18 embody the additional assumptions of unidimensionality and Axiom 3 over and above the general formulation of the choice model. Constraint equations to test these additional assumptions may be found by equating estimates of the u_{ij} from Equation 18 with estimates from Equation 10 to give:

$$\frac{p_{g,g+1} p_{g+h,g+h-1}}{p_{g,g+h} p_{g+h,g}} \prod_{i=g+2}^{i=g+h} \frac{p_{i-1,i-2} p_{i-1,i}}{(p_{i-1,i-1})^2} = 1. \quad (19)$$

For $h = 2$, Equation 19 becomes:

$$\begin{aligned} (p_{g,g+1}/p_{g,g+2})(p_{g+2,g+1}/p_{g+2,g}) \\ \cdot (p_{g+1,g} p_{g+1,g+2}/p_{g+1,g+1}^2) = 1, \end{aligned} \quad (20)$$

or, in additive form:

$$S_g = \ln p_{g,g+1} - \ln p_{g,g+2} + \ln p_{g+2,g+1} - \ln p_{g+2,g} - \ln p_{g+1,g} + \ln p_{g+1,g+2} - 2 \ln p_{g+1,g+1} = 0. \quad (21)$$

The S_g are denoted choice sensitivity critical parameters because they are derived from imposing constraints on the sensitivity parameters. Equations 19, like Equations 14, provide $\frac{1}{2}(n-1)(n-2)$ constraints. The reasoning which gives this number of constraints is precisely analogous to the reasoning which shows that Equations 14 also provide $\frac{1}{2}(n-1)(n-2)$ constraints. Equations 18 reduce the number of independent sensory parameters from $\frac{1}{2}(n-1)n$ to $n-1$, a reduction precisely of $\frac{1}{2}(n-1)(n-2)$, thus exhausting the possible constraints implied by Choice Axiom 3.

Each of Equations 17 have six and each of Equations 21 have eight probabilities estimated from the empirical data. Each empirically determined probability is subject to error. Obviously, the more terms in a constraint equation, the less accurately that equation can be tested, so that it is an advantage to manipulate Equations 16 and 20, if possible to produce equivalent sets of constraints which are functions of fewer empirical parameters. Such a simplification is achieved by multiplying Equations 16 by Equations 20 to provide Equations 22 and dividing Equations 20 by Equations 16 to provide Equations 24.

$$(p_{g+2,g+1}/p_{g+2,g}) \cdot (p_{g+1,g}/p_{g+1,g+1}) = 1, \quad (22)$$

or, in additive form:

$$C_g^- = \ln p_{g+2,g+1} - \ln p_{g+2,g} + \ln p_{g+1,g} - \ln p_{g+1,g+1} = 0 \quad (23)$$

and

$$(p_{g,g+1}/p_{g,g+2})(p_{g+1,g+2}/p_{g+1,g+1}) = 1, \quad (24)$$

or, in additive form:

$$C_g^+ = \ln p_{g,g+1} - \ln p_{g,g+2} + \ln p_{g+1,g+2} - \ln p_{g+1,g+1} = 0. \quad (25)$$

C_g^- and C_g^+ are choice critical test parameters which are derived from 2×2 submatrices composed of elements from the major diagonal and either elements 1 or 2 units below the diagonal (C^-) or elements 1 or 2 units above the diagonal (C^+). As will be shown later, the form of Equations 25 is very similar to that of the constraints produced by the logistic model so

that a juxtaposition with the equivalent logistic constraints will clarify the differences between the two models. Equations 17 and 21 still have the advantage that each tests specific predictions of the choice model.

The analysis of the choice model provides equations to estimate the parameters of the model and further equations to test the underlying axioms of the model. The next section performs a similar analysis of the Thurstonian model with a logistic probability distribution of difference between presented stimuli and internal criteria.

The Thurstone Model

The intuitions underlying the Thurstone model of categorical judgment are very well known. It is assumed that numerous presentations of a stimulus S_i gives rise to a distribution of effects, x on a psychological continuum. For stimulus S_i , $x = s_i + x_i$, where the s_i are fixed stimulus scale values and the x_i are independent identically distributed (i.i.d.) random variables with a common probability density function, f . The subject partitions the psychological continuum into m regions corresponding to the m responses by setting $m-1$ internal criteria, C_g . Analogously to the stimuli, the criteria also give rise to distributions, y , on the psychological continuum where, for criterion C_g , $y = c_g + y_i$, where the c_g are fixed criterion scale values and the y_i are i.i.d. random variables with a common probability density function f' . The basic theory does not require that f and f' be the same density functions or that the x_i be independent of the y_g . The subject makes a response less than or equal to g whenever $x - y \leq 0$. Then the probability of a response less than or equal to g to stimulus i , P_{ig} , is given by:

$$P_{ig} = \int_{-\infty}^0 k(s_i - c_g + t) dt = \int_{-\infty}^{-(s_i - c_g)} k(u) du, \quad (26)$$

where k is the probability distribution function of the difference between x_i and y_g . The P_{ig} are cumulative probabilities of the judgment matrix.

Obviously, f and f' uniquely determine k , but the converse is not true in general. The results of any category judgment experiment are completely specified by giving the P_{ig} for all $i = 1, 2, \dots, n$ and all $g = 1, 2, \dots, m-1$. Consequently, two models which have different distributions, f, f' , but give rise to the same difference distribution, k , are completely equivalent in any conceivable experiment. Hence, one may specify the type of Thurstone model by specifying the difference distribution k . Using this system of nomenclature, a logistic Thurstone model is one where the distribution k is logistic and similarly a normal Thurstone model has a normal distribution for k . Traditionally, the type of Thurstone model

is given by specifying the distributions f and f' (assumed identical). For the normal model, the two systems are equivalent, since, if k is normally distributed, f and f' must also be normally distributed. Adams and Messick (Note 1) have shown that, in simple discrimination or paired comparisons (i.e., $n = m = 2$), the choice model is equivalent to the logistic Thurstone model as defined here. Yellott (1977) has extended the result to multiple comparisons, i.e., where a subject picks that stimulus from among n stimuli which is highest on some psychological dimension. Then Yellott shows that the logistic Thurstone model for multiple comparison is completely equivalent to the choice model. He also shows that if the f are double exponential, then k is logistic; furthermore, if $n = 2$, there are definitely other distributions besides the double exponential which generate a logistic k , whereas if $n \geq 3$, then k logistic implies that the f are double exponential. (In paired comparison experiments, the difference distributions are between pairs of stimuli rather than between stimuli and criteria, so f is automatically the same as f' .) It will be shown that, for $n = m \geq 3$, the logistic Thurstone model of the absolute identification process is not equivalent to the choice model as it gives rise to different, and conflicting, constraint equations. The logistic Thurstone model is easier to work with than the normal Thurstone model as one can obtain the distribution function (cumulative density) and hence the constraint equations in closed form. However, the normal and logistic form are effectively indistinguishable experimentally because the distributions are so similar in shape (see Burke & Zinnes, 1965). It is important to note that the fact that the underlying stimulus and criterion distributions are such different shapes as the normal and double exponential does not imply that the predictions from the models will be very different. For it is in the nature of the experiment that the results depend only on the *difference* distributions. Similar difference distributions imply similar experimental results, no matter how different the underlying stimulus and criterion distributions. Consequently, evidence which favors the logistic Thurstone model over the choice model also favors the normal Thurstone case V model [where x_i, y_g are all normal (0,1)]. Furthermore, it is no use invoking the asymmetries in the double exponential to explain, for example, the anomalous high discriminability at the high end of a category scale. The difference distribution derived from the asymmetrical double exponentials is the symmetrical logistic, which actually generates the scale values.

Equation 26 enables one to calculate the stimulus and criterion scale values from the empirical cumulative probabilities provided that the form of k is known. In particular, when k is logistic, Equation 26 gives:

$$P_{ig} = \frac{1}{1 + \exp - (c_g - s_i)} \text{ for all } i, g. \tag{27}$$

Hence:

$$\exp - (c_g - d_i) = \frac{1 - P_{ig}}{P_{ig}} = A_{ig}. \tag{28}$$

A_{ig} is the odds in favor of a response less than or equal to g vs. a response greater than g . The A_{ig} determine the scale separation of stimulus i and response criterion g . To obtain the separation of criteria g and h , Equation 28 for criterion g is divided by Equation 28 for criterion h , with the stimulus i held constant:

$$\exp - (c_g - c_h) = \frac{A_{ig}}{A_{ih}}. \tag{29}$$

Similarly, the separation between stimulus i and stimulus j is given by:

$$\exp - (s_i - s_j) = \frac{A_{jg}}{A_{ig}}. \tag{30}$$

In fact, if Equations 29 specify criterion separation for adjacent criteria ($g = h + 1$), then additivity of the c_g implies all the Equations 29 for which $h - g > 1$. That is, there are only $m - 2$ independent criteria separations given by Equation 31, and similarly, $n - 1$ independent stimulus separations given by

$$\exp - (c_g - c_{g+1}) = A_{ig} / A_{i,g+1} \tag{31}$$

$$\exp - (s_i - s_{i+1}) = A_{i+1,g} / A_{ig}. \tag{32}$$

The most general form of the logistic Thurstone constraint equations may then be obtained by requiring that estimates of $c_g - c_{g+1}$ from adjacent rows of the judgment matrix be identical, giving:

$$\frac{A_{ig} \cdot A_{i+1,g+1}}{A_{i,g+1} \cdot A_{i+1,g}} = 1 \tag{33}$$

($g = 1, 2, \dots, n - 2$; $i = 1, 2, \dots, n - 1$). If one allows g to range from 1 to $m - 2$, corresponding to the $m - 2$ criteria separation, then Equations 33 comprise $(n - 1)(n - 2)$ constraints, which, together with the $n - 1$ independent s_i and $n - 1$ independent c_g , uses up all the df in the empirical judgment matrix. Equations 33 only use elements from the diagonal and one unit above and below it. The most sensitive test of the model comes from using elements on and near the diagonal, i.e., $g = i$ or $g = i - 1$; thus:

$$(A_{gg}/A_{g,g+1}) \cdot (A_{g+1,g+1}/A_{g+1,g}) = 1.$$

$$A_{ig} = A_{i+1,g} \text{ and } A_{i,g+1} = A_{i+1,g+1} \quad (39)$$

In additive form:

for $i = 2, g = 1$. Then, by induction over i, g from Equation 33, one can show that Equation 39 holds for all i, g . Equations 39, which hold if both the choice and the logistic Thurstone model are true, imply that the cumulative odds and hence the complete judgment matrix is contingent on the response alone, a situation which occurs only when the stimuli are completely indistinguishable. Thus, the two models are equivalent only in the limiting case where no discrimination between pairs of stimuli is possible.

$$L_g^o = (\ln A_{gg} - \ln A_{g,g+1}) + (\ln A_{g+1,g+1} - \ln A_{g+1,g}) = 0 \quad (34)$$

Of course, one can only apply constraint equations if $n > 2$. It is well known (Luce, 1963) that if $n = 2$, then the choice model and the logistic Thurstone model are equivalent in the sense that they give equivalent distance measures for the psychological separation of the two stimuli. In our terms, the distance measure of the choice model, d_C , is $-\ln u_{12}$; then, taking logarithms of Equation 10, one obtains

and:

$$(A_{g+1,g}/A_{g+1,g+1}) \cdot (A_{g+2,g+1}/A_{g+2,g}) = 1,$$

or in additive form:

$$L_g^- = (\ln A_{g+1,g} - \ln A_{g+1,g+1}) + (\ln A_{g+2,g+1} - \ln A_{g+2,g}) = 0 \quad (35)$$

$$g = 1, 2, \dots, n-2.$$

The L_g^o are logistic critical test parameters with cumulative odds up to and including the diagonal and positions adjacent to the diagonal. The L_g^- are logistic critical test parameters with cumulative odds up to the diagonal and positions either one or two units below the diagonal. Equations 34 and 35 test the logistic Thurstone model in a similar way to that which Equations 23 and 25 test the choice model.

$$d_C = \frac{1}{2}(\ln p_{11} + \ln p_{22} - \ln p_{12} - \ln p_{21}).$$

For the logistic Thurstone model, the distance measure, d_L , is equal to $s_2 - s_1$, and taking logarithms of Equation 30, one obtains:

$$d_L = [-\ln P_{21} + \ln(1 - P_{21}) - \ln(1 - P_{11}) + \ln P_{11}]$$

Comparison of Logistic Thurstone and Choice Models

for $g = 1$, as there is only one criterion in a two-choice experiment. For the two-choice experiment, $P_{11} = p_{11}, 1 - P_{11} = p_{12}, P_{21} = p_{21}, 1 - P_{21} = p_{22}$. Hence, $d_L = 2d_C$. The multiplicative factor of 2 does not destroy the equivalence, since any theorem which requires additivity of the d_L will automatically ensure additivity of the d_C , and vice versa.

Once one has obtained constraint equations for the two superficially similar models, one may investigate whether the two sets of constraints conflict. It will be shown that the set of choice constraints embodied in Equations 22 conflict with the set of logistic Thurstone constraints embodied in Equations 33.

It can now be seen that, in general, the logistic Thurstone model of category scaling is not equivalent to the choice model, although if $n = m = 2$, one gets the well-established equivalence of the two models for simple discrimination. Comparison of Equations 10 (choice) and Equations 30 (logistic) show that, in general, choice model formulations use simple probabilities, whereas Thurstone model formulations use cumulative probabilities. Obviously, if $n = 2$, simple and cumulative probabilities are the same, and hence the equivalence.

Let $g = 1$ and let $p_{21} = a, p_{22} = b, p_{31} = c$, and $p_{32} = d$. From Equation 22, one obtains

$$(b/a)(c/d) = 1, \quad (36)$$

and using 35 with 28, the defining equation of the A_{ig} , gives:

$$\frac{(c+d)(1-(a+b))a(1-c)}{(a+b)(1-(c+d))c(1-a)} = 1. \quad (37)$$

EMPIRICAL COMPARISON OF MODELS

If there is no conflict, then 36 and 37 will both hold, and from 36, $(c+d)/(a+b) = c/a$ may be substituted in 37, which, with some rearrangement, leads to $(1-c)/(1-a) = d/b$, which, combined with Equation 36, gives:

$$(1-c)/(1-a) = c/a. \quad (38)$$

Equation 38 can only be true if $a = c$, and hence $b = d$. Since $a + b = c + d$, one has also that

Having established that the two models make different theoretical predictions, a category judgment task must be arranged to produce the largest possible difference in experimental results between the two models. Violations of models are then detected either by using the constraint equation approach or by using optimization techniques. In the experiment described here, the task is an eight-stimulus, eight-

response category judgment of loudness. There are two conditions of bias, denoted unbiased (neutral) and biased (where subjects are encouraged to systematically underestimate all stimuli). Originally, the two conditions of bias were chosen with a view to showing that Thurstone scale values in the biased condition were a linear transform of the scale values in the unbiased condition (Jacobs, 1974). For the present purposes, this was a fortunate choice, since it turns out that the difference in predictions of the two models is considerably larger for the asymmetric judgment matrix generated by the biased condition than for the symmetric matrix of the unbiased condition.

Before describing the particular experiment, the two methods of comparing models will be discussed in some detail since each method has its own peculiar advantages and problems.

The Constraint Equation Method

The constraint equations are so formulated that, if they are violated, some critical test parameter calculated from the empirical probabilities will be different from zero. Since the observed critical test parameter is unlikely to be exactly zero, one needs some method of estimating the expected error in the critical parameter and hence evaluating if the deviation from zero is statistically significant. The additive (i.e., logarithmic) form of the constraint equations, 17, 21, 23, 25, 34, and 35, have been chosen with this problem in mind.

First, some preliminaries about errors of observation and their propagation. If some test parameter, T, is a function of n observed quantities, x_i, i = 1, 2, ..., n, then the error in T, ΔT, is given by:

$$(\Delta T)^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta x_i^2 + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n r_{ij} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \Delta x_i \Delta x_j \quad (40)$$

where Δx_i is the error in x_i, r_{ij} is the correlation between x_i, x_j (see, for example, Martin, 1971). From 40, one can deduce the important result for a function of one variable that

$$\Delta(1nx) = \Delta x/x, \quad (41)$$

if x is a probability, p, estimated from N attempts, then for the binomial distribution,

$$\Delta(1n p) = \{(1-p)/Np\}^{1/2}. \quad (42)$$

Furthermore, if probabilities come from different rows of the judgment matrix, they are uncorrelated, r = 0. If pairs of probabilities come from the same row, then simple probabilities are negatively correlated with each other and cumulative probabilities are

positively correlated with each other. Of course, the magnitude of r is not known, but we will calculate the error by using that value of r which would produce the maximum propagated error.

For the choice model, all constraints are of the form:

$$T = x + y + z,$$

so that

$$(\Delta T)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \quad (43)$$

where x,y,z, are independent, as they come from different rows of the judgment matrix. For Equations 23 and 25, there is no z and x,y are both of the form x = 1n p_{ig} - 1n p_{ih}. Then the maximum possible error in x occurs when the correlation between p_{ig} and p_{ih} is -1 and is given by:

$$(\Delta x)^2 = [\Delta(1n p_{ig}) + \Delta(1n p_{ih})]^2. \quad (44)$$

Equation 17, for the choice sensitivity criterion, has z of the same form as x and y. However, Equation 21 has z of the form:

$$z = 1n p_{ig} + 1n p_{ih} - 21n p_{ik}.$$

Since the correlations must all be negative, the maximum error occurs when the r_{ig,ih} = 0 and r_{ig,ik} = r_{ih,ik} = -1, and is given by:

$$(\Delta z)^2 = [\Delta(1n p_{ig}) + \Delta(1n p_{ik})]^2 + [\Delta(1n p_{ih}) + \Delta(1n p_{ik})]^2. \quad (45)$$

Use of Equations 42, 43, 44, and 45 in conjunction with the appropriate constraint equation enables one to estimate the SD of all the choice critical test parameters. A z-score is then calculated as the observed value of the critical parameter divided by the estimated SD. Then one can test whether the observed z-value is significantly different from zero, at any desired level of confidence. A modest 10% confidence level (two-tailed because the direction of violation is not known) is chosen since the method of derivation of the SD is known to be an overestimate.

For the logistic Thurstone model, the estimation is slightly more complicated. All constraints are of the form: T = u + v, where u,v are independent, since they come from different rows, so that:

$$(\Delta T)^2 = (\Delta u)^2 = (\Delta v)^2 \quad (46)$$

and u,v are both of the form:

$$u = 1n(1 - P_{ig}) - 1n P_{ig} + 1n P_{ih} - 1n(1 - P_{ig}).$$

Applying Equation 40 and noting that maximum error occurs when the correlation between P_{ig} and P_{ih} is zero, one obtains for the maximum error in u :

$$(\Delta u)^2 = (\Delta P_{ig})^2/[P_{ig}(1 - P_{ig})]^2 + (\Delta P_{ih})^2/[P_{ih}(1 - P_{ih})]^2. \quad (47)$$

Then use of Equations 42, 46, and 47 in conjunction with the logistic constraint equations, 34 and 35, enables one to estimate the SD and equivalent z -value of the logistic critical parameters.

Using these methods, the choice and logistic models are tested in ways which are closely analogous. With an 8×8 judgment matrix, Equations 23 and 25 give 12 independent tests of the choice model, and Equations 34 and 35 give 12 independent tests of the logistic model, for each judgment matrix. One may also calculate the mean and SD of the six critical test parameters embodied in each of the sets of Equations 23, 25, 34, and 35 and then calculate the t -value with 5 df and test its significance. The calculation of an empirical SD for each set of constraint equations also provides a rough check on the theoretical estimates of SD via Equation 40. As a very crude estimate, one would expect that:

$$SD_{\text{theory}} = (1 + r)^{1/2}SD_{\text{observed}}, \quad (48)$$

where r is the mean correlation between probabilities from the same row. Equation 48 can do no more than check that the theoretical estimates are of the right order when a model is not violated. If the model is violated, the argument breaks down as there is no reason to believe that all the test parameters have the same nonzero value.

A further problem, which applies both to the individual criterion tests of z -scores and to the group t tests described above, is the problem of power. The power of the test of a critical parameter depends on how accurately it can be estimated. So if model A is violated and not model B but the power of the test for A is larger, then one might hesitate to name B the better model. It is always possible that if B could be tested with equal power, it would also violate its constraints. The problem is relevant here since the tests of the choice model, dependent on simple probabilities, are slightly more powerful than the tests of the logistic, based on cumulative probabilities (the cumulative probabilities are more often close to zero or one and hence have higher fractional errors). Thus, the comparisons described above favor the logistic model because its constraints are tested with lower power. On the other hand, the comparison technique described below, using an optimum separator, favor the model tested with higher power, in this case the choice model.

If one chooses an arbitrary separator value, s , and

counts a violation when the critical test parameter $> s$ and no violation when the critical test parameter $\leq s$, then one can compare two models (or two sets of constraint equations) according to the split between numbers of violators and numbers of nonviolators [the chi-squared contingency test is used when the total number of tests is over 24; the Fisher-Yates test (Siegel, 1956, p. 96) is used for total number equal to 24]. An optimum value of the separator s is chosen to make the most difference between the sets of constraints being tested. If s is very large, there are no violations for either set of constraints; if s is small, there are no nonviolators for either set. The optimum value of s is easily found by inspection. This optimum separator test favors the high-power-low-SD model, since it is less likely to produce violations by chance. In fact, the expected number of violations, v_e , may be calculated; it is simply equal to the number of constraints tested multiplied by $(1 - \text{power of test})$. An estimate of the average power of the test for a set of constraint equations is determined from the z -score in the usual way, where z -score is S_{optimum} divided by mean theoretical SD, m_t , for the set of constraints. If one uses the estimate of maximum theoretical SD as described above, one of course overestimates the SD, hence underestimates the power and overestimates v_e . A more realistic estimate of v_e is obtained if all theoretical estimates of SD are reduced by a factor equal to the overall mean observed SD divided by the overall mean theoretical SD. The procedure is somewhat dubious, since it depends on the *observed* SD. If one used the observed SD direct, one should, overall, automatically come out with v_e close to the observed number of violations, v_o . Using binomial probability estimates, one may also estimate the probability of a value of $|v_e - v_o| \geq$ to the observed value.

If both the sets of tests which favor the model tested most powerfully and the sets which favor the model tested least powerfully end up favoring the same model, then one can be convinced that the constraint comparison is both fair and powerful.

A major advantage of the constraint method is that it will identify which of several assumptions in a model is at fault. In particular, investigation of the choice constraints in Equations 17 and 21 enable us to identify the source of the violation of choice theory axioms, even though Equations 23 and 25 provide a more powerful test of the overall model. In the present study, the constraint equation method is used to compare two models. It is also useful in investigating a single model with a view to accepting, rejecting, or indeed modifying contingent on which precise constraint was violated.

Optimization Methods

Optimization methods are used for two purposes: first, to determine those values of the theoretical

parameters of a model which give the "best fit" to the empirical data; and second, to determine whether the difference between the predictions of the model (using the best values of the theoretical parameters) and the observed data are so large as to invalidate the model, or whether the differences are of a magnitude which could have arisen by chance. "Best fit" implies some criterion for measuring the difference between the data predicted by the model and the observed data and then ensuring that this difference is as small as possible. It is well known that there are several possible criteria (least squares, minimum chi-squared, maximum likelihood, etc.). One is not obliged to use the same criterion to evaluate the best fit and to test the goodness of fit. However, it does seem desirable wherever this is technically feasible, and for this reason the minimum chi-squared criterion has been chosen for this study.

In order to find the values of the theoretical parameters, x_t , which minimize a criterion C , it is of course necessary to be able to express C as a function of the x_t . One also has to be able to evaluate $\delta C/\delta x_t$ for all the x_t , but this may be done numerically, if the analytic function is too complicated. C , the chi-squared statistic, is defined by Equation 49.

$$C = \sum_{i=1}^8 \sum_{g=1}^8 \frac{(f_{ig} - o_{ig})^2}{f_{ig}}, \quad (49)$$

where o_{ig} is the observed frequency with which response g is given to stimulus i and f_{ij} is the frequency of response g to stimulus i predicted by the model in question. The f_{ij} are functions of the x_t which depend on the model, where of course

$$f_{ig} = N_i p_{ig} \quad (50)$$

if stimulus i is presented N_i times.

All the models considered here have 14 theoretical parameters (seven sensory and seven motivational for an eight-stimulus, eight-response category judgment task). The normal Thurstone model is considered as well as the logistic Thurstone model as the optimization methods do not require that the theoretical parameters be expressed in terms of the probabilities in closed form.

For the choice model, the theoretical parameters, x_t , are the seven motivational parameters, b_g ($g = 2, 3, \dots, 8$, where b_1 is set to unity for normalization purposes), and the seven sensory parameters, $u_{i,i+1}$ ($i = 1, 2, \dots, 7$). So the p_{ig} , and hence the f_{ig} , may be calculated from the b_g and the $u_{i,i+1}$ from Equations 1, 2, 10, and 11.

For the Thurstone models, the theoretical parameters are the seven criterion positions, c_g ($g = 1, 2, \dots, 7$) and the seven psychological stimulus

positions, s_i ($i = 2, 3, \dots, 8$, where s_1 is set to zero as the arbitrary start point of an interval scale). The simple probabilities, p_{ig} , may be calculated from the cumulative probabilities which are directly predicted by the model:

$$P_{ig} = P_{i,g+1} - P_{ig}, \quad (51)$$

$$i = 1, 2, \dots, 8; g = 1, 2, \dots, 7.$$

For the logistic model, the P_{ig} are given by Equation 27.

For the normal model, the P_{ig} are given by Equation 26, where $k(u)$ is the normal density function, $N(0,1)$.

The optimization problems which require solution are of the general form: given empirical observations o_k and $C = C(x_t, o_k)$, find the values of x_t which minimize C . The methods used here were developed by the Numerical Optimisation Centre, Hatfield Polytechnic, for solving the general problem, and are embodied in their FORTRAN programs OPVM and OPND1. OPVM finds an unconstrained local minimum for C using a quasi-Newton algorithm. Details are given by Biggs (1971, 1973). OPND1 is a routine which calculates numerical derivatives by the method of central differences.

The judgment matrices are not necessarily completely filled, i.e., some of the o_{ig} are zero. This is not in itself a difficulty, as it simply cuts down the number of df. Frequencies that are small but nonzero are more difficult, as one would like a solution which is not dominated by the high-variability/low-frequency cells. For this reason, the solution is based only on those cells where the observed frequency is greater than 5, with of course the consequent reduction in df. In general, the number of df is the number of cells used in the summation of Equation 49, minus 14 (for the estimated parameters) minus 8 (because the sum of frequencies in each row is fixed). One does not want to use the low-frequency cells to determine the x_t , but once the x_t are obtained, it is reasonable to ask if they also fit the omitted cells. So a chi-squared value based on the x_t with $o_{ig} > 5$ was calculated for all cells with predicted frequencies $\geq .5$.

EXPERIMENT

Method

Subjects. Two undergraduates from Columbia University served as subjects. They were selected from 10 students, who replied to an advertisement in the student newspaper, because they performed best on a 1-h training session in categorical judgment. Both subjects had normal audiograms. The subjects were paid approximately \$2/h. One dollar was a flat-rate minimum. The remaining earnings were contingent on the subject's performance. The subjects worked for points, gaining when correct and losing when incorrect, in a pattern depending on the bias condition. The points were then converted to dollars at a rate determined by pretesting to give earnings of roughly \$1/h.

Sessions. Each session was 1 h long and comprised 400 trials, 240 trials in 24 min followed by a break, followed by 160 trials in 16 min. Each subject performed on each condition until a predetermined number of trials had been collected from "asymptotic" sessions. "Asymptotic" sessions are consecutive sessions where the variability in percent correct is no more than would be expected by chance. For both subjects, the predetermined number of trials was 2,000 for the unbiased condition (250 presentations of each stimulus), i.e., five asymptotic sessions. In the biased conditions, D.P. performed six asymptotic sessions (300 presentations of each stimulus) and D.J. eight asymptotic sessions (400 presentations of each stimulus).

Stimuli. The stimuli were 1-sec bursts of 500 Hz pure tones, ranging in intensity in 1/2-dB steps from 71.1 dB (Stimulus 1) to 74.6 dB (Stimulus 8).

Responses. The response system was the keyboard of a "touch-tone" telephone with the letters and digits 0 and 9 painted out, so that the subject used the keys labeled 1-8 for Responses 1-8.

Outcome and feedback. The subjects were instructed that Response 1 was correct for the softest stimulus, Response 2 for the next softest, and so on up to Response 8, which was correct for the loudest stimulus. The subject was faced with a large outcome display which gave feedback and payoff information on each trial. Feedback was given by a digital lamp which showed the number corresponding to the stimulus just presented. Payoff was indicated by one of three lights: corresponding to a response less than the actual stimulus (underestimate), a correct response, or a response greater than the actual stimulus (overestimate). The payoff, i.e., the number of points earned, was indicated next to the appropriate light.

In the unbiased (neutral) condition, the subject earned 10 points for a correct response and lost 2 points for both over- and underestimates. In the biased condition, the subject still earned 10 points for a correct response, but he lost 9 points for an overestimate error and only 1 point for an underestimate error. Instructions emphasized that the subject's task was to gain as many points (and hence as much money) as possible, not necessarily to achieve the maximum number of correct responses.

Procedure. The subject was seated in a dental chair in a sound-attenuating booth (IAC 1203). The response system was on the right arm of the chair, facing the outcome display through the booth window. The apparatus was set to give trials comprising a 1-sec burst of tone followed by 5 sec of silence, during which the subject made a response and received payoff and feedback information and the experimenter selected the stimulus for the

next trial. On each trial, the stimulus presented and response given were recorded on paper tape for future analysis. Stimuli were presented in quasi-random order with the constraint that each stimulus was presented 10 times in each block of 80 trials.

The subjects were informed of the total number of points gained and equivalent money earned both at the midsession break and at the end of each session.

Results

Judgment frequency matrices for both subjects and both conditions are shown in Tables A and B of the appendix.

Constraint equations. Table 1 summarizes the results of testing all constraint equations for both subjects in the two conditions of bias. Each set of test equations comprises 6 tests for each judgment matrix. The results from the two subjects are pooled, making 12 tests of each set of constraints in each condition. In the unbiased condition, the power of the tests for L⁰ and L⁻ and also for C⁺ and C⁻ are very similar and hence have been pooled. In the biased condition, 2, the tests for L⁻ and C⁻ are considerably more powerful than for L⁰ and C⁺, so the results are shown separately. The first column simply gives N, the number of constraints tested.

The number of constraints, N_z, for which the z-score, based on the theoretical maximum SD, is greater than the critical z-value for the 10% confidence limit (two-tailed) and is shown in the second column of Table 1. For both the choice and the logistic model, there are 12 independent constraints per judgment matrix, i.e., 48 constraints in all. None of these constraints are violated for the logistic model. For the choice model, there are 2/24 violations in Condition 1 at the 10% level of confidence, certainly not enough to provide conclusive evidence against the choice model. In the biased condition, 2, the pic-

Table 1
Number of Violations of Constraint Equations with Different Criteria of Violation and t Values for Each Set of Constraints

Model	Set of Constraints	N	N _z	Sign† Comparison	N _c	Sign† Comparison	t Test		
							t _{D.P.}	t _{D.J.}	df
Condition 1									
LOG	L ⁰ , L ⁻	24	0}	n.s.	1}	.100	-2.17	-2.50*	11
CHO	C ⁺ , C ⁻	24	2}		7}		-1.69	2.99**	11
CHO	S	12	1}	n.s.	3}	n.s.	-.42	-.19	5
CHO	B	12	1}		3}		.71	2.62*	5
Condition 2									
LOG	L ⁰	12	0}	n.s.	1}	n.s.	.14	1.32*	5
CHO	C ⁺	12	1}		4}		2.35	2.60*	5
LOG	L ⁻	12	0}	.001	1}	.001	1.18	-2.69	5
CHO	C ⁻	12	12}		11}		6.66**	11.43**	5
LOG	L ⁰ , L ⁻	24	0}	.001	2}	.001	1.12	.55	11
CHO	C ⁺ , C ⁻	24	13}		15}		6.95**	6.61**	11
CHO	S	12	2}	n.s.	4}	.025	4.09**	-.98	5
CHO	B	12	3}		10}		4.94**	5.12**	5

Note—N = number tested; N_z = number with z ≥ 1.65; N_c = number with c > .41; c is any critical test parameter. *p > 5% **p > 1% †Significance level of probability that number of violations from bracketed sets of constraints are different. Fisher-Yates exact probability test where total number = 24. Chi-squared contingency test where total number > 24.

ture is quite different: all 12 of the powerful C⁻ constraints are violated at the 10% level, providing conclusive evidence against the choice model. Since there are two violations of the choice sensory constraints and three violations of the choice bias constraint, no conclusion can be drawn as to the source of the problem with the choice model.

The last three columns of Table 1 provide evidence based on the mean extent of violation of the constraint for a set of constraint equations. The t-values are shown separately for each subject and each set of constraints. The results in Condition 1 are somewhat equivocal. For D.P., neither model is disconfirmed, whereas for D.J., the t-value is significantly different from zero for both models—for the logistic model merely at the 5% level, while for the choice model

at the 1% level. In the biased condition, t-values for both subjects for C⁻ constraints and hence for the combined C⁺ and C⁻ constraints are highly significant. The only evidence against the logistic comes from D.J. L⁻ constraints, again at the weak 5% level.

Tables 2 and 3 give details of the exact magnitude of each constraint coefficient, and its theoretical maximum SD and associated z-score. The ratio of mean theoretical maximum SD, m_t, to the empirically estimated SD ranges from .9 to 2.6 with a mean of 1.55 corresponding to an approximate value of r = 0.24 from Equation 48. Thus, the theoretical estimates of SD seem "reasonable" when compared with empirical estimates.

The optimum separator method was used both to compare logistic and choice model constraints and,

Table 2
Magnitude of Constraint Coefficients and Theoretical Estimates of Maximum Errors
with Associated Z Scores for Unbiased Condition 1

Model	Equation	Parameter	g Value						Mean	Empirical SD
			1	2	3	4	5	6		
Subject D.P.										
Logistic	34	L ⁰	-.26	-.10	-.06	-.41	-.15	-.01	-.16	.15
		err	.27	.27	.28	.26	.27	.30	.28	
		Z	-.96	-.37	-.21	-1.58	-.55	-.03	-.16	.61
	35	L ⁻	-.23	-.30	.34	.10	.08	-.23	-.07	.24
		err	.29	.29	.29	.29	.27	.27	.39	
		Z	-.79	-1.03	1.17	.34	.30	-.85	-.14	.88
Choice	25	C ⁺	.38	.05	.06	-.65	.02	.27	.21	.36
		err	.41	.36	.45	.39	.33	.33	.38	
		Z	.92	.14	.13	-1.66*	.06	.82	.14	.87
	23	C ⁻	-.02	-.19	-.07	.48	.30	.01	.08	.25
		err	.40	.39	.41	.42	.38	.31	.39	
		Z	-.05	-.49	-.17	1.14	.79	.03	.21	.62
21	S	.38	.23	.13	-1.02	-.41	.14	-.09	.52	
	err	.45	.50	.55	.53	.47	.42	.49		
	Z	.84	.46	.24	-1.92*	-.87	.33	.16	.58	
17	B	.36	-.15	-.01	-.37	.31	.28	.08	.27	
	err	.58	.63	.72	.68	.61	.54	.63		
	Z	.62	-.24	-.01	-.40	.51	.52	.17	.44	
Subject D.J.										
Logistic	35	L ⁰	.20	-.03	-.43	-.19	.09	-.16	-.09	.22
		err	.28	.30	.28	.29	.28	.31	.35	
		Z	.71	-.10	-1.53	-.66	.32	-.52	-.30	.79
	34	L ⁻	.04	-.10	-.28	-.19	-.24	-.11	-.15	.11
		err	.31	.30	.30	.28	.27	.28	.29	
		Z	.13	-.33	-.93	-.68	-.89	-.39	-.52	.40
Choice	25	C ⁺	.47	-.02	.01	.21	.40	.19	.21	.20
		err	.33	.39	.38	.33	.34	.36	.36	
		Z	1.42	-.05	.03	.64	1.18	.53	.62	.59
	23	C ⁻	.74	.45	.03	-.04	-.18	.49	.25	.36
		err	.34	.38	.40	.39	.35	.34	.37	
		Z	2.18†	1.18	.08	-.10	-.51	1.44	.71	1.05
21	S	-.27	-.47	-.02	.26	.59	-.29	-.03	.40	
	err	.44	.52	.54	.49	.45	.47	.49		
	Z	-.61	-.90	-.04	.53	1.31	-.62	-.05	.84	
17	B	1.21	.42	.05	.17	.22	.69	.46	.43	
	err	.55	.63	.66	.62	.58	.58	.60		
	Z	2.20†	.67	.08	.27	.38	1.19	.80	.79	

*p > 10% †p > 5%

Table 3
Magnitude of Constraint Coefficients and Theoretical Estimates of Maximum Errors
with Associated Z Scores for Biased Condition 2

Model	Equation	Parameter	g Value						Mean	Empirical SD
			1	2	3	4	5	6		
Subject D.P.										
Logistic	35	L ⁰	.24	-.15	.03	-.16	-.04	.13	.01	.16
		err	.27	.29	.34	.31	.32	.42	.33	
		Z	.88	-.52	.09	-.52	-.13	.31	.02	.54
	34	L ⁻	.29	-.00	-.21	-.01	.23	.30	.10	.21
		err	.22	.23	.24	.23	.24	.28	.24	
		Z	1.32	-.01	-.88	-.04	.96	1.07	.40	.85
Choice	25	C ⁺	.62	.01	.22	.03	.11	.34	.22	.23
		err	.40	.47	.49	.46	.44	.54	.47	
		Z	1.55	.02	.45	.07	.25	.63	.49	.57
	23	C ⁻	1.02	.83	.47	.51	.81	.41	.68	.25
		err	.17	.21	.23	.24	.25	.23	.22	
		Z	6.00†	3.95†	2.05**	2.13**	3.24**	1.78*	3.19	1.61
17	B	-.39	-.83	-.25	-.48	-.69	-.07	-.45	.27	
	err	.39	.44	.53	.50	.47	.54	.48		
	Z	-1.00	-1.89*	-.47	-.96	-1.47	-.13	-.98	.64	
21	S	1.64	.83	.69	.38	.76	.74	.84	.42	
	err	.48	.53	.62	.61	.58	.66	.58		
	Z	3.42†	1.57	1.11	.62	1.31	1.12	1.53	.98	
Subject D.J.										
Logistic	35	L ⁰	.46	.03	.38	-.29	.27	.03	.15	.28
		err	.39	.43	.50	.40	.38	.51	.43	
		Z	1.18	.07	.76	-.73	.71	.06	.34	.68
	34	L ⁻	-.25	-.21	-.35	-.02	-.02	-.51	.22	.20
		err	.28	.28	.27	.27	.25	.34	.28	
		Z	-.89	-.75	-1.30	-.07	-.08	-1.50	-.74	.64
Choice	25	C ⁺	.91	.41	.23	-.13	.46	.25	.36	.34
		err	.52	.60	.66	.58	.51	.65	.59	
		Z	1.75*	.73	.35	-.22	.90	.38	.65	.66
	23	C ⁻	.77	.51	.51	.44	.51	.61	.56	.12
		err	.21	.25	.26	.27	.27	.26	.25	
		Z	3.67†	2.04**	1.96**	1.66*	1.89*	2.35**	2.26	.73
21	S	.13	-.08	.28	-.99	-.05	-.37	-.18	.45	
	err	.44	.52	.54	.49	.45	.47	.49		
	Z	.29	-.15	.52	-2.02**	-.11	-.79	-.38	.92	
17	B	1.68	.96	.75	.31	.98	.86	.92	.44	
	err	.55	.63	.66	.62	.58	.58	.60		
	Z	3.05**	1.52	1.13	.50	1.69*	1.48	1.56	.84	

*p > 10% **p > 5% †p > 1%

within the choice model, to compare the sensory constraints with the bias constraints. Column 4 of Table 1 shows N_C , the number of constraint coefficients, C , greater than the optimum separator value, s . The value of s was found to be 0.41 by inspection. The fifth column of Table 1 indicates the significance level of the comparison of N_C s from the bracketed sets of conditions. For Condition 1, N_C for the choice model is significantly larger than N_C for the logistic model only at the weak 10% level. For Condition 2, the difference between C^- and L^- is, as usual, highly significant and is, of course, the major cause of the highly significant difference between the combined choice constraints and combined logistic constraints. In Condition 2, the number of choice

bias violations is also significantly greater than the number of choice sensory violations.

Since $s_{optimum}$ is less than the 10% critical value for many of the critical constraint coefficients, one would expect some violations using this separator value. Table 4 shows the predicted number of violations, v_e , using both the mean theoretical SD for each set of constraints, m_t , as an estimate of the SD of the constraint coefficient, and more realistically, $m_t/1.55$. The power estimate on which the v_e estimate is made is also shown. One may also calculate the binomial probability of a difference $|v_o - v_e|$ greater than or equal to that obtained in the experiment, as shown in Table 4. Using the low-power theoretical maximum SD, one may conclude that

Table 4
 Number of Constraint Coefficients Greater than Optimum Separator Value, v_o , with Number of Predicted Violations, v_e , and Power of Test Based on Different Theoretical Estimates of SD of Constraint Coefficients

Model	Constraint	Number Tested	v_o	Estimate of SD = Theoretical Max.			Estimate of SD = Theoretical Max./1.55		
				v_e	Power	Significance $ v_o - v_e $	v_e	Power	Significance $ v_o - v_e $
Condition 1									
LOG	L^o, L^-	24	1	4.0	.83	n.s.	.8	.97	n.s.
CHO	C^+, C^-	24	7	6.5	.73	n.s.	2.1	.91	4×10^{-3}
CHO	S	12	3	4.8	.60	n.s.	2.2	.81	n.s.
CHO	B	12	3	6.0	.50	n.s.	3.5	.71	n.s.
Condition 2									
LOG	L^o	12	1	3.4	.72	n.s.	1.1	.91	n.s.
CHO	C^+	12	4	5.0	.58	n.s.	2.5	.79	n.s.
LOG	L^-	12	1	1.3	.89	n.s.	.2	.99	n.s.
CHO	C^-	12	11	1.0	.92	9×10^{-11}	.1	.99	1×10^{-21}
LOG	L^o, L^-	24	2	4.7			1.3		
CHO	C^+, C^-	24	15	6.0			2.6		
CHO	S	12	4	4.8	.60	n.s.	2.2	.81	n.s.
CHO	B	12	10	5.9	.51	.04	3.3	.72	1×10^{-4}

11/12 C^- violations are unlikely to have arisen by accident given the power of the test for $s = 0.41$ (.99). Similarly, the number of choice bias violations is significant at the 5% level. If one uses the less conservative estimate of the SD, one may also conclude that the combined number of choice violations in the unbiased condition is significant.

In summary, by whatever criterion one chooses, there can be no doubt that the C^- choice constraints in the biased condition are not obeyed. There is also considerable evidence that it is the choice bias, rather than the choice sensory constraints, which is at fault. There is some evidence that the choice constraints are violated in the unbiased condition. For subject D.J. only, there is weak evidence against the logistic model in the form of significant t-values for the combined logistic constraints in Condition 1 and the L^- constraints in Condition 2.

Optimization method. Table 5 shows the minimum chi-squared values for the best fitting solutions to choice, normal Thurstone, and logistic Thurstone

models. Only cells of the judgment matrix with frequencies greater than 5 were used to estimate the optimum theoretical parameters. All the chi-squared values for both types of Thurstone model are less than the critical chi-squared value for the 5% confidence level. For Condition 1, the choice chi-squared values are significant at the .5% level for both subjects. The probability of such high values occurring by chance is in the region of 1:1,000. In Condition 2, the choice chi-squared values are too high for there to have been even the remotest probability of their occurring by chance. As predicted in the theory section, there is little to choose between the normal and logistic version of the Thurstone model.

Having chosen the theoretical parameters to fit the cells with frequencies greater than 5, it is interesting to know if these values also give a reasonable fit to the low-frequency cells. With this in mind, the chi-squared values based on these theoretical parameters were calculated using all cells with expected frequencies greater than or equal to .5. As may be seen from

Table 5
 Chi-Squared Values for Best-Fitting Solutions from Normal and Logistic Thurstone Models and Choice Model

	Condition 1—Unbiased						Condition 2—Biased					
	D.P.			D.J.			D.P.			D.J.		
	Choice	Logistic	Normal	Choice	Logistic	Normal	Choice	Logistic	Normal	Choice	Logistic	Normal
Parameter A												
χ^2	56.9	36.7*	40.4*	52.6	29.1*	31.6**	289.0	21.9*	32.1**	245.5	19.6*	15.5*
df	30	30	30	21	21	21	23	23	23	20	20	20
Parameter B												
χ^2		50.8*	45.2*		46.4*	46.9**		35.1*	47.5†		34.9*	25.5*
df		42	40		36	33		38	32		35	29

Note—Parameter A: Minimum χ^2 from all cells where observed frequency > 5. Parameter B: χ^2 calculated with same parameter values as A but including all cells where expected frequency $\geq .5$.
 *measured $\chi^2 < \chi^2$ 10% **measured $\chi^2 < \chi^2$ 5% †measured $\chi^2 < \chi^2$ 2.5%

row B of Table 5, even with this stringent condition, there is no evidence against either version of the Thurstone model.

DISCUSSION

The results of the present study provide strong evidence against the choice model of category judgment. Conversely, Thurstone models, whether based on a logistic or a normal difference distribution, make an impressive fit to the empirical judgment matrices.

With hindsight, it appears that the reason the choice formulation is unsuccessful is intimately bound up with the fact that the stimuli are actually ordered psychologically along a single continuum. This unidimensional ordering is automatically built into the Thurstone models. However, if one examines the matrix of Equation 5 for the choice model, one can see that it would be no different if the stimuli were not ordered, but just happened to have the relations between discriminabilities expressed in Equation 18. In other words, there is no natural way of formulating the choice model of categorical judgment so that the stimuli and responses must be ordered along a

single continuum.

It is interesting to note that the very strong conclusions one can draw about both choice and Thurstone models stem from the fact that an asymmetric biased condition was included in the experimental design.

Both the constraint method and the optimization method provided useful tests of the two models. In general, the optimization method was most discriminating, since it was able to provide conclusive evidence against the choice model even in the unbiased condition. However, the constraint-equation approach had the advantage of showing that the problem with the choice model was fundamental, since it arose from the bias constraints which depend directly on the axioms, not just on the sensory constraints which arise from specific constraints on the discriminabilities (Equation 18). If one is testing a new model for the first time, one can usually get a good idea of whether it is satisfactory from a quick look at the constraint equations, although one may have to go to the more powerful optimization methods to prove one's conclusion. Obviously, one great practical advantage of the constraint method is that it doesn't require a computer and associated technology.

APPENDIX

Table A
Judgment Frequency Matrix for Both Subjects in the Neutral Condition

Stimulus	Subject D.P.									Subject D.J.									
	Response								Total	Response								Total	
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		
1	133	56	33	19	6	3	1		251	105	94	33	13	5				1	251
2	59	71	61	34	13	8	5		251	81	89	50	24	5	4				253
3	39	46	72	42	23	14	13	1	250	27	62	81	38	25	16	3	2		254
4	10	31	40	55	32	50	29	7	254	7	25	51	75	50	34	8	1		251
5	2	13	32	41	55	45	45	16	249		8	25	38	75	63	35	9		253
6	1	6	9	23	45	67	68	36	255		5	11	27	51	71	59	28		252
7		6	6	19	28	56	82	57	254	1	1	1	14	44	51	89	51		252
8			2	9	16	49	72	103	251				1	23	29	83	115		251
Total	244	229	255	242	218	292	315	220	2015	221	284	252	230	278	268	277	207		2017

Table B
Judgment Frequency Matrices for Both Subjects in the Biased Condition

Stimulus	Subject D.P.									Subject D.J.									
	Response								Total	Response								Total	
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		
1	293	87	20	6	2	1			409	227	64	8	3						302
2	199	126	54	22	6		1		408	168	97	30	7	2					304
3	91	159	100	41	12	4			407	84	105	80	29	6					304
4	21	89	128	104	38	16	6	1	403	23	76	97	73	19	11	2	1		302
5	4	44	85	111	92	40	19	2	397	3	34	67	84	69	35	12			304
6	4	14	36	87	120	86	46	10	403		8	28	69	88	68	37	6		304
7		2	11	52	84	135	85	26	395	1	2	13	31	72	93	77	16		305
8			5	21	52	127	120	75	402			4	10	24	73	111	80		302
Total	612	523	439	444	406	409	271	114	3224	506	386	327	306	280	280	239	103		2427

REFERENCE NOTE

1. Adams, E., & Messick, S. *An axiomatization of Thurstone's successive intervals and paired comparison scaling models*. Technical Report 12, Applied Mathematics and Statistics Laboratory, Stanford University, 1957.

REFERENCES

- ARBUCKLE, J., & NUGENT, J. A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical & Statistical Psychology*, 1973, **26**, 240-260.
- BURKE, C. J., & ZINNES, J. L. A paired comparison of paired comparisons. *Journal of Mathematical Psychology*, 1965, **2**, 53-76.
- BIGGS, M. Minimisation algorithms making use of nonquadratic properties of the objective function. *Journal of the Institute of Mathematical Applications*, 1971, **8**, 315-327.
- BIGGS, M. A note on minimisation algorithms which make use of the nonquadratic properties of the objective function. *Journal of the Institute of Mathematical Applications*, 1973, **12**, 337-338.
- EDWARDS, W., & TVERSKY, A. (Eds.), *Decision making*. London: Penguin, 1967.
- GULLICKSEN, H. A least squares solution to successive intervals assuming unequal variance. *Psychometrika*, 1954, **19**, 117-139.
- JACOBS (KORNBROT), D. E. *Separation of response bias and stimulus effects in category judgment experiments*. Unpublished doctoral dissertation, Columbia University, 1974.
- LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- LUCE, R. D. Detection and recognition. In R. D. Luce, R. Bush, & E. H. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
- LUCE, R. D., & GALANTER, E. H. Psychophysical scaling. In R. D. Luce, R. Bush, & E. H. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
- LUCE, R. D., & SUPPES, P. Preference, utility and subjective probability. In R. D. Luce, R. Bush, & E. H. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3). New York: Wiley, 1965.
- MARTIN, B. K. *Statistics for physicists*. London: Academic Press, 1971.
- SHEPARD, R. N. Stimulus response generalization. *Psychometrika*, 1957, **22**, 325-345.
- SIEGEL, S. *Nonparametric statistics for the behavioural sciences*. London: McGraw-Hill, 1956.
- THURSTONE, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, **38**, 368-389.
- YELLOTT, J. I. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment and the double exponential distribution. *Journal of Mathematical Psychology*, 1977, **15**, 109-146.

(Received for publication November 21, 1977;
revision accepted July 14, 1978.)