

# A Probabilistic IRT Model for Unfolding Preference Data

David Andrich  
Murdoch University

A probabilistic model is developed for the pair-comparison design in which the unfolding principle that governs the choice process uses a discriminial process analogous to Thurstone's Law of Comparative Judgment. However, this process is governed by the square of the distance between the location of the person and the stimulus, rather than controlled by the lo-

cation of the stimulus as in Thurstone's formulation. A simulation study demonstrates the feasibility of estimation, and two examples use real data to show the implications of the unfolding models for psychological research. *Index terms: choice data, item response theory, pair comparisons, preference data, unfolding.*

In his major statement on the measurement of attitude, Thurstone (1928) discussed two related but distinct designs with different response processes; one set pertained to scale construction and the other to person measurement. In scale construction, the basic data collection design required a *pair-comparison response* and the response process modeled according to the Law of Comparative Judgment (LCJ; Thurstone, 1927). This approach allows a set of statements to be mapped onto an idealized linear continuum in which both order and interstatement distances are meaningful.

One way to ascertain the model's accuracy is to determine how closely the distances  $d_{AB}$  between statements  $A$  and  $B$ ,  $d_{BC}$  between  $B$  and  $C$ , and  $d_{AC}$  between  $A$  and  $C$  conform to the relationship  $d_{AC} = d_{AB} + d_{BC}$ , where the location order of the statements is  $A < B < C$ . Because of the time-consuming nature of the pair-comparison design, other data collection designs, such as rank ordering (Thurstone, 1931) and equal-appearing intervals (Thurstone & Chave, 1929), were developed with corresponding modifications to the LCJ. Nevertheless the pair-comparison design remains fundamental in theoretical psychometrics.

The LCJ for pair comparisons assumes that the response process is *cumulative* in that stimuli further along a continuum are more likely to be perceived as such, relative to the other stimuli, by *every person*. As a consequence, the location of the individuals making the comparisons is *assumed* not to affect the choice of one stimulus over another. As stated by Thurstone (1959),

If the scale is to be regarded as valid, the scale values of the statements should not be affected by the opinions of the people who help to construct it. This may turn out to be a severe test in practice,

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 13, No. 2, June 1989, pp. 193-216  
© Copyright 1989 Applied Psychological Measurement Inc.  
0146-6216/89/020193-24\$2.45

but the scaling method must stand such a test before it can be accepted as being more than a description of the people who construct the scale. (p. 228)

Although Thurstone never formalized his own verbal rationale for the LCJ in a way that showed that the location of the person making the comparison is eliminated, this has been done (Andrich, 1978a).

In the person-measurement stage the data collection design involved a *direct response*: Persons agreed or disagreed with each statement, and the measurement was the median or mean of the scale values of the statements endorsed. In contrast to the formal LCJ, the person-measurement procedure, though plausible, is essentially ad hoc. However, in addition to being less rigorous, the procedure reflects a response process that differs from Thurstone's pair-comparison design. The process assumes the person's location by hypothesizing that the person will tend to agree with those statements that have scale values similar to that of the person, and will tend to disagree with statements with scale values far away from those of the person at *either* extreme. These tendencies provide the basis for the *unfolding* response process.

The cumulative and unfolding principles are not confined to the pair-comparison and the direct-response designs, respectively—the cumulative principle can apply to direct responses to statements and the unfolding principle can apply to pair comparisons between statements. The former gives the well-documented Guttman scale and its probabilistic counterpart, Rasch's simple logistic model (Andrich, 1985a; Rasch, 1960/1980), while the latter has been studied extensively by Coombs and his colleagues (Coombs, 1950; Coombs & Avrunin, 1977; Coombs & Smith, 1973).

Coombs' original work with pair comparisons assumed that the persons and statements or other stimuli were located on a unidimensional continuum, and the choice of one stimulus over another was governed by the distances  $d_{PA}$  and  $d_{PB}$  between person  $P$  and stimuli  $A$  and  $B$  respectively: If  $d_{PA} < d_{PB}$  then the person was more likely to select  $A$  than  $B$ . Ramsay (1980) distinguished the cumulative and the unfolding principles by considering the former to have a preferred direction and the latter to have a preferred point. The preferred point reflected the position of the person. It was assumed that the person perceived the relative locations of the statements in the same way, and that the differences in choices of preferences were due only to differences among the locations of the persons.

To maintain a clear distinction between the cumulative and unfolding mechanisms in a general pair-comparison design, the term *pair comparison* will refer to the cumulative case where attributes are compared in items, irrespective of a person's location. The term *pairwise preference* will refer to the unfolding case where the person's location governs the choice with respect to the attribute.

Coombs' original methods for identifying the locations of the persons and the stimuli were deterministic; no error term was formalized in the response or choice process. Substantial work has been done since Coombs' first conceptions of the unfolding process in pairwise preference designs. Explicit probabilistic models have been formulated (Sixtl, 1973), extensions into more than one dimension have been made (DeSarbo & Hoffman, 1986; Schönemann, 1970), and the possibility that persons not only select differently but perceive the relative locations of stimuli differently has been considered (Carroll, 1972). Tanzer (1983) related various approaches to unfolding.

This paper returns the discussion of the unfolding process for pairwise preferences to the first principles enunciated by Thurstone and followed by Coombs. These principles are reformulated to incorporate recent developments in person/item response theory (PIRT). Combining new developments with a reexamination of the first principles allows the derivation of a simple and tractable probabilistic model for unfolding choice or pairwise preference data in a single dimension. The model is a special case of that proposed by Schönemann and Wang (1972), Cooper and Nakanishi (1983), and Poole (1984), but deriving it from first principles clearly establishes the relationship between the model and the response mechanism.

The model is derived from two approaches. The first follows an approach used in PIRT where pairwise comparisons are implied from direct responses, and the second follows Thurstone's original derivation

of the LCJ. The model accounts for the two most salient factors which might be expected intuitively to govern preferences; one is the difference between the location of the person and the midpoint of the locations of the two stimuli, and the other is the difference between the locations of the two stimuli.

Conceptualizations of data which subscribe to the unfolding mechanism, and their analyses, have tended to be confined to literature in mathematical psychology. In applied psychological and educational measurement, the unfolding principles have been neglected, especially as cumulative PIRT has been emphasized in the last two decades. However, developmental, attitude, and choice data are not necessarily cumulative. The model and analyses considered in this paper permit dealing with unfolding unidimensional data in a manner essentially as straightforward as the present applications of PIRT models.

### The Probabilistic Unfolding Model for Pair Comparisons

Before deriving the proposed model from some plausible assumptions, it is necessary to review the main requirements of an unfolding model for pairwise preferences.

Figure 1, which shows the location of three stimuli, the midpoints between adjacent stimuli, and the locations of six persons, illustrates the assumed response mechanism. According to the unfolding principle, Persons 1 and 2 would tend to select Stimulus 1 over all others, and so on.

If a person were at the midpoint between two stimuli, then the person would tend to be just as likely to select one stimulus as the other. In addition, if two stimuli happened to be located at the same point, then irrespective of the person's location, the person should again be just as likely to select one stimulus as the other. All models of this process should satisfy these two particular "boundary" conditions.

#### A Modified Discriminal Process

Thurstone formalized a latent continuous discriminial process when person  $n$  encountered stimulus  $i$  according to

$$d_{ni}^* = \delta_i + e_{ni}^* \quad , \quad (1)$$

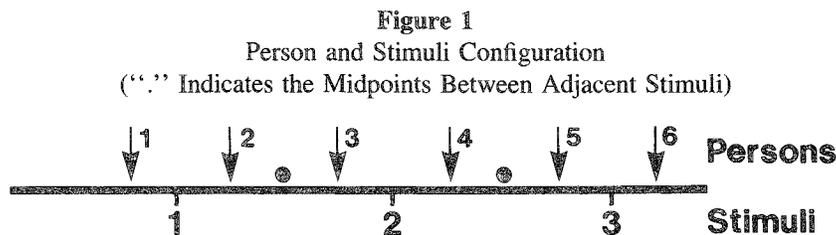
where

$$E(d_i^*) = \delta_i \quad (2)$$

and

$$\text{Var}(d_i^*) = \sigma_i^{*2} \quad . \quad (3)$$

The form of the expectation and variance indicates that the person is selected randomly from a population. The first modification to the specification of the discriminial process in Equations 1 through 3, which leads to an unfolding response process, is to resolve the error component  $e_{ni}^*$  so that it explicitly includes



a location parameter  $\beta_n$  for each person. This may be done as

$$e_{ni}^* = -\beta_n + e_{ni} \quad , \quad (4)$$

giving

$$d_{ni}^* = \delta_i - \beta_n + e_{ni} \quad . \quad (5)$$

In the resolution of the error, the opposite of  $\beta_n$  is chosen for convenience and without loss of generality (Andrich, 1978a). This discriminial process may be rewritten as

$$d_{ni} = -(\beta_n - \delta_i) + e_{ni} \quad , \quad (6)$$

with

$$E(d_{ni}) = -(\beta_n - \delta_i) \quad (7)$$

and

$$\text{Var}(d_{ni}) = \sigma_{ni}^2 \quad . \quad (8)$$

In Equation 7, the expectation explicitly involves the location of person  $n$ , with the error now entirely within the person. Formalizing the location of the person is essential when deriving an unfolding model based on a discriminial process. In addition, because each person is identified throughout, it is unnecessary to assume any random sampling of persons.

The second modification, similar to that made by Davison (1977), assumes that the value of the discriminial process is not governed simply by the difference in locations between the person and the stimulus, but by the square of this difference, thus rendering the actual direction of the difference irrelevant. Such a modification (Andrich, 1988) may be written as

$$d_{ni} = -(\beta_n - \delta_i)^2 + e_{ni} \quad . \quad (9)$$

This model assumes that the greater the distance between  $\beta_n$  and  $\delta_i$ , the less likely the person is to select stimulus  $i$ , and all factors being equal, the less likely the person is to select, prefer, or endorse stimulus  $i$  over stimulus  $j$ . Suppose then that person  $n$ , if required to endorse stimulus  $i$ , would do so if  $d_{ni} > 0$ . Clearly, as the distance between  $\beta_n$  and  $\delta_i$  increases, the person is less likely to endorse it.

The probability of endorsement or nonendorsement is governed not only by the locations  $\beta_n$  and  $\delta_i$ , but also by the distribution of the error  $e_{ni}$ . Thurstone assumed that this random error distribution was normal. If the same assumption is made regarding Equation 9, then the probabilities of endorsement and nonendorsement, designated respectively by  $p_{ni1}$  and  $p_{ni0}$ , are given by

$$p_{ni1} = \frac{1}{(2\pi)^{1/2}\sigma_{ni}} \int_0^\infty \exp\left\{-\frac{1}{2}\left[\frac{t - (\beta_n - \delta_i)^2}{\sigma_{ni}}\right]^2\right\} dt \quad (10)$$

and

$$p_{ni0} = 1 - p_{ni1} \quad . \quad (11)$$

Assuming a homogeneous error variance  $\sigma_{ni}^2 = \sigma^2$  across all persons with respect to all stimuli permits this common term to be absorbed into the location parameters as an arbitrary unit.

Because it is numerically equivalent (Johnson & Kotz, 1972), yet more tractable, it has become customary in modern psychometric theory to replace the cumulative normal distribution with the logistic. The logistic distribution is obtained by assuming that the error distribution is the double exponential rather than the normal (Yellott, 1977). If the substitution is made, and assuming homogeneous within-person error variance, then the respective probabilities simplify to

$$p_{ni1} = \frac{1}{\eta_{ni}} \exp[-(\beta_n - \delta_i)^2] \quad (12)$$

and

$$p_{ni0} = \frac{1}{\eta_{ni}}, \quad (13)$$

where

$$\eta_{ni} = 1 + \exp[-(\beta_n - \delta_i)^2] \quad (14)$$

is a normalizing factor. Clearly, as the distance between  $\beta_n$  and  $\delta_i$  increases, the probability of an approved response decreases.

### A Conditional Motivation for a Probabilistic Model

There are four straightforward possible outcomes when the person encounters both stimuli: endorsement of both, nonendorsement of both, or endorsement of  $i$  and nonendorsement of  $j$  and vice versa. On the assumption of stochastic independence between these instantaneous outcomes, their probabilities are shown in Table 1.

Although the instantaneous outcomes include four possibilities, the task for the person is choosing between the two stimuli. From outcomes (0,0) and (1,1), no distinction exists between the two stimuli, while from the outcomes (0,1) or (1,0), the choice follows immediately.

Suppose that if person  $n$  is disposed initially to an outcome which does not distinguish between the stimuli [(0,0) or (1,1)], that the person is required to reconsider those reactions. A plausible, though nontestable, assumption is that the distribution of these unacceptable original outcomes (0,0) and (1,1) among the acceptable ones, (1,0) and (0,1), is such that the latter retain their original relative probabilities.

The probability of the outcome (1,0) or (0,1), then, which implies a choice of stimulus  $i$  over stimulus  $j$  or vice versa, is obtained by normalizing the probability within the outcome set [(1,0) or (0,1)]. The probability that person  $n$  selects stimulus  $i$  over stimulus  $j$  is given, on normalization and simplification, by

$$p_{nij} = \frac{p(1,0)}{p[(1,0) \text{ or } (0,1)]} = \frac{\exp[-(\beta_n - \delta_i)^2]}{\exp[-(\beta_n - \delta_i)^2] + \exp[-(\beta_n - \delta_j)^2]} \quad (15)$$

Further straightforward simplifications reveal the desirable properties suggested earlier for such a model. First, Equation 15 may be rearranged to

Table 1  
Possible Outcomes for Two Variables  $i$  and  $j$   
and Probabilities Assuming Stochastic Independence

$i$	$j$	Probability
0	0	$\frac{1}{\eta_{ni}\eta_{nj}}$
1	0	$\frac{1}{\eta_{ni}\eta_{nj}} \exp[-(\beta_n - \delta_i)^2]$
0	1	$\frac{1}{\eta_{ni}\eta_{nj}} \exp[-(\beta_n - \delta_j)^2]$
1	1	$\frac{1}{\eta_{ni}\eta_{nj}} \exp[-(\beta_n - \delta_i)^2 - (\beta_n - \delta_j)^2]$

$$p_{nij} = \frac{1}{\gamma_{nij}} \exp[-(\beta_n - \delta_i)^2 + (\beta_n - \delta_j)^2] \quad , \quad (16)$$

where

$$\gamma_{nij} = 1 + \exp[-(\beta_n - \delta_i)^2 + (\beta_n - \delta_j)^2] \quad (17)$$

is a normalizing factor.

Second, the exponent in Equation 16 may be simplified readily from

$$-(\beta_n - \delta_i)^2 + (\beta_n - \delta_j)^2 \quad (18)$$

to

$$2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2) \quad , \quad (19)$$

giving

$$p_{nij} = \frac{1}{\gamma_{nij}} \exp[2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2)] \quad (20)$$

where

$$\gamma_{nij} = 1 + \exp[2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2)] \quad . \quad (21)$$

Equation 20 is the form of the model most appropriate for estimation. However, a third simplification of the exponent from

$$\begin{aligned} 2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2) &= 2\beta_n(\delta_i - \delta_j) - (\delta_i - \delta_j)(\delta_i + \delta_j) = (\delta_i - \delta_j)[2\beta_n - (\delta_i + \delta_j)] \\ &= 2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij}) \quad , \end{aligned} \quad (22)$$

giving

$$p_{nij} = \frac{1}{\gamma_{nij}} \exp[2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij})] \quad , \quad (23)$$

where  $\bar{\delta}_{ij} = (\delta_i + \delta_j)/2$  is the midpoint between  $\delta_i$  and  $\delta_j$ , reveals explicitly the features of this model which govern the choice of one stimulus over another.

Equation 23 clearly shows that the probability of selecting stimulus  $i$  over stimulus  $j$  depends on the distance between  $\delta_i$  and  $\delta_j$  given by the factor  $(\delta_i - \delta_j)$ , and on the distance of  $\beta_n$  from the midpoint between  $\delta_i$  and  $\delta_j$  given by the factor  $(\beta_n - \bar{\delta}_{ij})$ . The desirable properties suggested earlier for a model are evidently met in that when  $\delta_i = \delta_j$  or when  $\beta_n = \bar{\delta}_{ij}$ , the probability of the choice of  $i$  rather than  $j$  is .5. Further, the closer  $\beta_n$  is to  $\delta_i$  relative to  $\delta_j$ , the more likely the person is to select  $i$  rather than  $j$ .

Clearly,

$$p_{nji} = 1 - p_{nij} = \frac{1}{\gamma_{nji}} \exp[2(\delta_j - \delta_i)(\beta_n - \bar{\delta}_{ij})] \quad . \quad (24)$$

### Relationship to the Bradley-Terry-Luce Model

The above argument parallels that presented in Andrich (1978a) in which it was shown that the simple logistic response model (SLM) of Rasch (1960/1980, 1961) has a special relationship to Thurstone's Case V of the LCJ (Thurstone, 1927). It was shown that with a suitable introduction of a person parameter as in Equation 1 of the discriminial process, and its subsequent elimination in forming comparisons, the equations for pair comparisons associated with the SLM are equivalent to the equations associated with Case V of the LCJ. Specifically, and through the elimination of the person parameters, both derivations

lead to the Bradley-Terry-Luce (BTL) model for pair comparisons (Bradley & Terry, 1952; Luce, 1959) which takes the form

$$p_{ij} = \frac{\exp(\delta_i - \delta_j)}{1 + \exp(\delta_i - \delta_j)} \quad (25)$$

where  $p_{ij}$  is the probability that stimulus  $i$ , with location parameter  $\delta_i$ , will be chosen over stimulus  $j$  with location parameter  $\delta_j$ . This model is appropriate for cumulative response data and includes no person parameter. From both the rationale leading to the model of Equation 23 and from the structure of the model, it is clear that this model is the unfolding counterpart to the BTL model for a cumulative response process. In particular, the factor  $(\delta_i - \delta_j)$  appears in both models, with the unfolding model having a second factor,  $(\beta_n - \bar{\delta}_{ij})$ , containing the location of the person. Luce (1977) and Yellott (1977) provided other connections between Thurstone's law of comparative judgment and the BTL model, with Luce reexamining Thurstone's discriminial process. Coombs (1964) also noted a relationship between the BTL model and Rasch's SLM.

### A LCJ for Pair Comparisons

The relationships between Case V of the LCJ and the SLM for cumulative response data suggest that the model of Equation 16 may also be obtained from the perspective of the LCJ. This is indeed the case.

The motivation analogous to that of the LCJ commences again with the discriminial process of Equation 9. Considering discriminial processes with respect to stimuli  $i$  and  $j$  and expanding gives

$$d_{ni} = -\beta_n^2 - \delta_i^2 + 2\beta_n\delta_i + e_{ni} \quad (26)$$

and

$$d_{nj} = -\beta_n^2 - \delta_j^2 + 2\beta_n\delta_j + e_{nj} \quad (27)$$

The difference process when person  $n$  chooses between  $i$  and  $j$  is then given by

$$\begin{aligned} d_{nij} &= d_{ni} - d_{nj} = 2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2) + e_{ni} - e_{nj} = (\delta_i - \delta_j)[2\beta_n - (\delta_i + \delta_j)] + e_{ni} - e_{nj} \\ &= 2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij}) + e_{ni} - e_{nj} \end{aligned} \quad (28)$$

with

$$E(d_{nij}) = \mu_{nij} = 2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij}) \quad (29)$$

and

$$\text{Var}(d_{nij}) = \sigma_{nij}^2 = \sigma_{ni}^2 + \sigma_{nj}^2 \quad (30)$$

Note that no covariance term is involved in the expression for  $\text{Var}(d_{nij})$ . This follows from the specification that the stochastic part of the process is within the individual. As shown in Andrich (1978a), the covariance in the difference process for Thurstone's formulation of the LCJ arises from the interperson variation, and it is also eliminated when the person parameter is made explicit.

If  $d_{ni}$  and  $d_{nj}$  are normally distributed, it follows that  $d_{nij}$  is too, and the probability that stimulus  $i$  is selected over  $j$  corresponds to the probability that  $d_{nij} > 0$ . This probability  $p_{nij}$  is thus given by the cumulative normal:

$$p_{nij} = \frac{1}{(2\pi)^{1/2}\sigma_{nij}} \int_0^\infty \exp\left[-\frac{1}{2}\left(\frac{t - \mu_{nij}}{\sigma_{nij}}\right)^2\right] dt \quad (31)$$

By analogy to Case V of the LCJ, if homogeneous error variances  $\sigma_{ni}^2 = \sigma_{nj}^2 = \sigma^2$  are assumed, the unit  $\sigma$  may be absorbed into the location parameters, giving

$$p_{nij} = \frac{1}{(2\pi)^{1/2}} \int_0^\infty \exp\left[-\frac{1}{2}(t - \mu_{nij})^2\right] dt \quad (32)$$

Substituting once more the explicit logistic for the implicit cumulative normal function gives

$$p_{nij} = \frac{1}{\gamma_{nij}} \exp[2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij})] \quad (33)$$

where

$$\gamma_{nij} = 1 + \exp[2(\delta_i - \delta_j)(\beta_n - \bar{\delta}_{ij})] \quad (34)$$

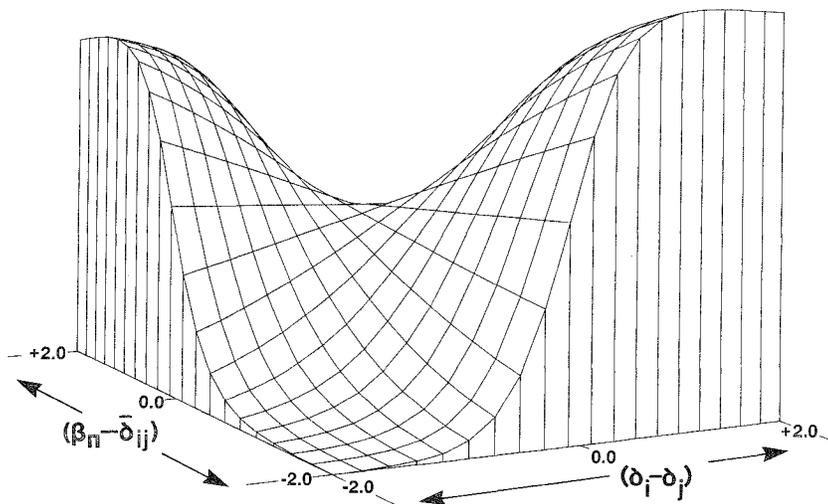
This expression is identical to Equation 23 derived by the earlier conditional argument and thus the two approaches produce the same model.

### Further Issues in the Preference of *i* Over *j*

Figure 2 shows the probability  $p_{nij}$  as a function of  $(\delta_i - \delta_j)$  and  $(\beta_n - \bar{\delta}_{ij})$ . It is clear from Figure 2, as well as Equations 16, 20, and 23, that as the difference  $\delta_i - \delta_j$  increases, the discrimination in the choice between *i* and *j* increases. Analogously, as  $\beta_n$  moves away from  $\bar{\delta}_{ij}$ , where  $\beta_n$  is closer to  $\delta_i$  than  $\delta_j$ , then the choice of *i* over *j* becomes stronger. This suggests that as a person's location gets further away from the location of the stimuli, the person becomes more *sensitive* to differences between the stimuli. The range in which this relationship holds is an empirical matter, and it may not hold as well at extremes. It seems plausible, however, for moderate values. For example, if a person had to choose between two foods which differed only in sweetness to a given degree, then according to the model, persons with a very sweet tooth (e.g.,  $\beta >> \bar{\delta}_{ij}$ ) would be more likely to select the sweeter of the two foods than persons with not such a sweet tooth. Thus the sweeter the tooth, the more important a particular

Figure 2

The Preference of *i* Over *j* as a Function of  $\beta_n$ ,  $\delta_i$ , and  $\delta_j$



difference in sweetness. Conversely, persons with a “sour tooth” ( $\beta < \bar{\delta}_{ij}$ ) would be more likely to select the less sweet of the two foods than persons without such a sour tooth.

This effect contrasts with Weber’s law in which the noticeable differences in objects are directly proportional to the values of the objects on the relevant variable. That is, according to Weber’s law, the greater the values of the objects, such as weight, the greater the differences required in order that they be noticed. However, Weber’s law deals explicitly with a cumulative, not an unfolding, mechanism, and within a range that is comfortable for the judge. Weber’s law would break down with weight lifters who need to “compare” weights some distance from their ideal point, that is, when the objects are extremely heavy (or extremely light) for them.

### Relating Applications of the Cumulative and Unfolding Mechanism

Comparison of the cumulative and unfolding models will proceed by referring to Jansen (1984). Jansen’s paper raised issues that illustrate the types of misunderstandings that arise when the distinctions between the cumulative and unfolding mechanisms are not retained at every level of consideration. Jansen’s starting point was a particular relationship that was described by Andrich (1978a) between Case V of Thurstone’s LCJ and Rasch’s SLM.

In that paper, Andrich translated Thurstone’s verbal description of a discriminial process directly into Equation 6, rather than Thurstone’s original formulation (Equation 1). The distinction between the two equations is that Equation 6 contains a person parameter while Equation 1 does not. The consequence of using Equation 6 rather than Equation 1 is that when Thurstone’s LCJ is applied for two items, the location parameter of person  $n$ ,  $\beta_n$ , is eliminated by subtraction, as the following example shows.

Consider a second stimulus  $j$ . Then

$$d_{nj} = \delta_i - \beta_n + e_{nj} \quad , \quad (35)$$

and the difference process is given by

$$d_{ij} = d_{ni} - d_{nj} = (\delta_i - \beta_n) + e_{ni} - (\delta_j - \beta_n) - e_{nj} = (\delta_i - \delta_j) + (e_{ni} - e_{nj}) \quad (36)$$

which simplifies to

$$d_{ij} = \delta_{ij} + e_{nij} \quad , \quad (37)$$

which does not contain  $\beta_n$ .

The elimination of the person parameter  $\beta_n$  when the stimulus parameters are compared provides the connection to the SLM. Moreover, when the logistic function is substituted for the numerically equivalent cumulative normal distribution, the equation for comparing two stimuli is identical (Equation 25).

In the SLM, the discriminial process from Equation 6, and by analogy to Equation 12, gives

$$p_{ni1} = \frac{1}{\gamma_{ni}} \exp[-(\beta_n - \delta_i)] \quad , \quad (38)$$

where  $p_{ni1}$  corresponds to the probability that person  $n$  answers item  $i$  in the negative direction. The response to the stimulus here is direct. (The reversal of response to convention is used in order to be consistent with Equation 6 where the  $\beta_n$  and  $\delta_i$  have signs opposite to the usual.)

Using the same probabilistic theory discussed previously with respect to Equation 15 gives

$$p_{nij} = \frac{p(1,0)}{p[(1,0) \text{ or } (0,1)]} = \frac{\exp(\delta_i - \delta_j)}{1 + \exp(\delta_i - \delta_j)} \quad , \quad (39)$$

which is identical to Equation 25, which is the BTL model.

The substantive counterpart to the mathematical identity of Equations 25 and 39, provided by Andrich (1978a) and acknowledged by Jansen, is that in scoring 0 on only one stimulus and 1 on the other, the ordering of the stimuli, as far as that person is concerned, can be inferred. This is because in scoring 1 on stimulus  $i$ ,  $\beta_n$  is deemed below  $\delta_i$ , and in scoring 0 on stimulus  $j$ ,  $\beta_n$  is deemed above  $\delta_j$ ; hence  $\delta_i$  is deemed above  $\delta_j$ .

It is important to stress that the equivalence of the above relationships rests on the fact that Equation 1 represents a cumulative mechanism, and thus Equations 25 and 39 must also refer to a cumulative mechanism.

Following an appreciation of the above relationship, Jansen proceeded to what he termed a “data-theoretical perspective.” This led him into *pairwise preference* data rather than *pair-comparison* data, and inevitably into probabilistic *unfolding* mechanisms. He then concluded that the equivalence set up by Andrich only follows when the locations of the person parameters are all on one side of the locations of the stimuli, that is, when they are unilateral. It certainly is the case that when all stimuli are on one side of the location of the person, the cumulative and unfolding mechanisms cannot be distinguished in data. However, Jansen implied that this result provides some kind of restriction to the relationship presented in Andrich (1978a). The restriction arises because a cumulative mechanism is considered in one type of response process (the pair comparison) and an unfolding (pairwise preference) in the other. *If the mechanism for both the direct response and the pair comparison is a cumulative one, then the parallel between the Thurstone and Rasch models described above is complete and unrestricted by locations of persons.*

Although the parallel between the direct and pair-comparison designs is complete for the cumulative mechanism, the circumstances under which the cumulative or unfolding mechanism would operate need to be examined (Andrich, 1985b). Consider an example. First, suppose that some achievement items have been constructed for assessing students and that they are designed to fit the SLM. Then the response to the items by the students will be cumulative, and it is important that the locations  $\beta_n$  of the students are *within* the range of the items. Otherwise, if student locations are substantially above or below the item difficulties, all responses will be either 0 or 1, and no comparison between items could be estimated.

Second, suppose that teachers who are expert in the subject matter are asked to compare the items in pairs, and to indicate which of the two items will be the more difficult (for students). Here it would be expected, and required, that the locations of the teachers is well above that of the difficulties of the items in order for them to make such a judgment accurately. Although it is an empirical matter whether the difficulties so obtained are equal to those obtained from the direct responses from a sample of students, the response process would be expected to be cumulative, not unfolding, when the teachers compare the items.

Suppose, thirdly, that judges are asked to compare the affective values of statements used to construct an attitude scale. Now the locations of judges need not be at the extremes of the locations of the statements—they can be within the range of the statements. However, the importance of the instruction to the judges needs to be appreciated. If the instruction is to make a comparison *irrespective of their own attitudes*, as required by Thurstone, then again the response will be of the cumulative kind. That their attitude should not affect the scalings is reflected in Equations 25 and 39. Suppose, finally, that the instruction is to indicate which of each pair of statements better reflects their own attitudes: Then the response will be of the unfolding kind. In that case the locations of the persons will affect the choice, and this is reflected in Equations 16, 20, and 23. Jansen (1984) did not make this distinction, but implied that the pair-comparison design must involve the unfolding mechanism.

Following the scaling of attitude statements with an instruction requiring the *cumulative* mechanism, Thurstone asked the same people whose attitudes were to be assessed to agree or disagree with the statements—that is, to respond directly to the statements. Here he expected the unfolding mechanism to

operate: To estimate the location of a person he obtained the mean of the values to which a person agreed. A formal discussion of the model of unfolding for direct responses is provided in Andrich (1988). Thus both cumulative and unfolding models are possible for both designs where objects are first compared and for designs that require direct responses: The case that is appropriate depends on the nature of the task and the types of stimuli.

### Maximum Likelihood Estimation

The most direct method for obtaining solution equations is that of maximum likelihood. Suppose the data are summarized as in Table 2 in which the cell entry  $f_{ij}$  indicates the number of times that stimulus  $i$  is chosen over stimulus  $j$ . This is the usual summary of pair-comparison and pairwise preference data. However, because the parameter of each person is involved explicitly, it is necessary to consider the results of the pairwise preference for each person. The choice of  $i$  over  $j$  by person  $n$  may be designated, as shown in Table 3, by the Bernoulli random variable  $X_{nij}$  which takes on the values  $x_{nij} = 1$  if  $i$  is chosen over  $j$ , and 0 otherwise.

In deriving the estimation equations, the form of the model expressed in Equation 20 will be used. With the introduction of the random variable  $X_{nij}$ , the probability that person  $n$  selects stimulus  $i$  over stimulus  $j$  is given by

$$p_{nij} = \frac{1}{\gamma_{nij}} \exp\{x_{nij}[2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2)]\} \quad (40)$$

where

$$\gamma_{nij} = 1 + \exp[2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2)] \quad (41)$$

In deriving the solution equations, it will prove efficient to note that

$$\frac{\partial(\log \gamma_{nij})}{\partial \beta_n} = 2(\delta_i - \delta_j)p_{nij} \quad (42)$$

and likewise that

$$\frac{\partial(\log \gamma_{nij})}{\partial \delta_i} = 2(\beta_n - \delta_i)p_{nij} \quad (43)$$

Table 2  
Number of Times Stimulus  $i$   
Is Chosen Over Stimulus  $j$

Stimulus	Stimulus						Total
	1	2	3 ... j ... J				
1	$f_{11}$	$f_{12}$	$f_{13}$	$f_{1j}$	$f_{1J}$	$f_{1+}$	
2	$f_{21}$	$f_{22}$	$f_{23}$	$f_{2j}$	$f_{2J}$	$f_{2+}$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
$i$	$f_{i1}$	$f_{i2}$	$f_{i3}$	$f_{ij}$	$f_{iJ}$	$f_{i+}$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
$I$	$f_{I1}$	$f_{I2}$	$f_{I3}$	$f_{Ij}$	$f_{IJ}$	$f_{I+}$	

Table 3  
 Results of Pair Comparisons  
 for a Single Person

Stimulus	Stimulus					Total
	1	2	3 ... j ... J			
1	$x_{11}$	$x_{12}$	$x_{13}$	$x_{1j}$	$x_{1J}$	$x_{1+}$
2	$x_{21}$	$x_{22}$	$x_{23}$	$x_{2j}$	$x_{2J}$	$x_{2+}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{ij}$	$x_{iJ}$	$x_{i+}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	$x_{I1}$	$x_{I2}$	$x_{I3}$	$x_{Ij}$	$x_{IJ}$	$x_{I+}$

**Joint Maximum Likelihood Estimates**

The probability or likelihood of the entire set of observations then is given by

$$L = \prod_n \prod_j \prod_i \frac{1}{\gamma_{nij}} \exp\{x_{nij}[2\beta_n(\delta_i - \delta_j) - (\delta_i^2 - \delta_j^2)]\} \quad (44)$$

Rearranging and combining terms turns out to be much simpler after taking the logarithm of Equation 38 and obtaining the required derivatives. Thus

$$\frac{\partial \log L}{\partial \beta_n} = 2 \sum_i \sum_j x_{nij}(\delta_i - \delta_j) - 2 \sum_i \sum_j (\delta_i - \delta_j) p_{nij} \quad (45)$$

Because the  $\delta_i$  and  $\delta_j$  refer to the one set of parameters, and because  $p_{nij} = 1 - p_{nji}$ , the further simplification that focuses on each stimulus  $i$  is

$$\frac{\partial \log L}{\partial \beta_n} = 2 \left( \sum_i s_{ni} \delta_i - \sum_i \delta_i \sum_{j \neq i} p_{nij} \right) \quad (46)$$

where

$$s_{ni} = \sum_{j \neq i} x_{nij} \quad (47)$$

may be carried out.

Analogously,

$$\frac{\partial \log L}{\partial \delta_i} = 2 \left( -T_i \delta_i + \sum_n s_{ni} \beta_n + \delta_i \sum_n \sum_{j \neq i} p_{nij} - \sum_n \beta_n \sum_{j \neq i} p_{nij} \right) \quad (48)$$

where

$$T_i = \sum_n \sum_{j \neq i} x_{nij} \quad (49)$$

One constraint, effectively the origin, is necessary to solve the above equations. This may be set so that the estimated stimulus locations sum to 0. Thus the final set of solution equations is given by

$$2 \left( \sum_i s_{ni} \hat{\delta}_i - \sum_i \hat{\delta}_i \sum_{j \neq i} \hat{p}_{nij} \right) = 0 \quad (n = 1, \dots, N) \quad (50)$$

$$2 \left[ \hat{\delta}_i \left( -T_i + \sum_n \sum_{j \neq i} \hat{p}_{nij} \right) + \sum_n \hat{\beta}_n \left( s_{ni} - \sum_{j \neq i} \hat{p}_{nij} \right) \right] = 0 \quad (i = 1, \dots, I) \quad (51)$$

and

$$\sum_i \hat{\delta}_i = 0 \quad (52)$$

### An Example of Solution Equations

Although the above equations are straightforward, examination of a simple example will show how the data are summarized. Consider again Figure 1 and suppose that the choice each of the six persons makes with respect to the three stimuli is exactly as expected. Then the choice matrix for each person is shown in Table 4, which also shows the choice matrix summed across persons. The marginals on the matrices, which are relevant to the solution equations, are also shown.

The solution equation for Person 1 then is

$$2\{2\hat{\delta}_1 + 1\hat{\delta}_2 + 0\hat{\delta}_3 - [\hat{\delta}_1(\hat{p}_{112} + \hat{p}_{113}) + \hat{\delta}_2(\hat{p}_{121} + \hat{p}_{123}) + \hat{\delta}_3(\hat{p}_{131} + \hat{p}_{132})]\} = 0 \quad (53)$$

and the solution equation for Stimulus 1 is

$$\begin{aligned} 2[\hat{\delta}_1(-5 + \hat{p}_{112} + \hat{p}_{113} + \hat{p}_{212} + \hat{p}_{213} + \hat{p}_{312} + \hat{p}_{313} + \hat{p}_{412} + \hat{p}_{413} + \hat{p}_{512} + \hat{p}_{513} + \hat{p}_{612} + \hat{p}_{613}) \\ + \hat{\beta}_1(2 - \hat{p}_{112} - \hat{p}_{113}) + \hat{\beta}_2(2 - \hat{p}_{212} - \hat{p}_{213}) + \hat{\beta}_3(1 - \hat{p}_{312} - \hat{p}_{313}) \\ + \hat{\beta}_4(0 - \hat{p}_{412} - \hat{p}_{413}) + \hat{\beta}_5(0 - \hat{p}_{512} - \hat{p}_{513}) + \hat{\beta}_6(0 - \hat{p}_{612} - \hat{p}_{613})] = 0 \quad (54) \end{aligned}$$

### A Solution Algorithm

The solution equations clearly require only the marginals  $s_{ni}$  ( $i = 1, \dots, I; n = 1, \dots, N$ ),  $r_n$  ( $n = 1, \dots, N$ ), and  $T_i$  ( $i = 1, \dots, I$ ). This is relatively convenient for estimation, though it requires a vector for each person with respect to each item.

The equations are solved iteratively using the Newton-Raphson algorithm. Although the solution involves multiple parameters, the univariate approach used in PIRT seems to be the most efficient, primarily because the matrix of second derivatives, which includes a parameter for each person, becomes prohibitively large for inversion if a multivariate approach is used. This procedure is similar to that employed by Lord (1968), Bock (1972), Jensema (1974), Wright and Douglas (1977), and Andrich (1978b) with models in PIRT where person and item parameters appear simultaneously.

Table 4  
Choice Matrices of Six Persons With Respect to Three Items

Person 1					Person 2					Person 3									
$j$					$j$					$j$									
$i$	1	2	3	$s_{1i}$	$i$	1	2	3	$s_{2i}$	$i$	1	2	3	$s_{3i}$					
1		1	1	2	1		1	1	2	1			1	1					
2	0		1	1	2	0		1	1	2	1		1	2					
3	0	0		0	3	0	0		0	3	0	0		0					
Person 4					Person 5					Person 6					Total Choices				
$j$					$j$					$j$					$j$				
$i$	1	2	3	$s_{4i}$	$i$	1	2	3	$s_{5i}$	$i$	1	2	3	$s_{6i}$	$i$	1	2	3	$T_i$
1		0	0	0	1		0	0	0	1		0	0	0	1		2	3	5
2	1		1	2	2	1		0	1	2	1		0	1	2	4		4	8
3	1	0		1	3	1	1		2	3	1	1		2	3	3	2		5

Thus defining

$$\phi(\beta_n) = 2 \left( \sum_i s_{ni} \delta_i - \sum_i \delta_i \sum_{j, j \neq i} p_{nij} \right) \quad (n = 1, \dots, N) \quad (55)$$

and

$$\psi(\delta_i) = 2 \left[ \delta_i \left( -T_i + \sum_n \sum_{j, j \neq i} p_{nij} \right) + \sum_n \beta_n \left( s_{ni} - \sum_{j, j \neq i} p_{nij} \right) \right] \quad (i = 1, \dots, I) \quad (56)$$

the solution algorithm for  $\phi(\beta_n) = 0$  and  $\psi(\delta_i) = 0$  is given by

$$\hat{\beta}_n^{(q+1)} = \beta_n^{(q)} - \left[ \frac{\phi(\beta_n)}{\phi'(\beta_n)} \right]^{(q)} \quad (57)$$

and

$$\hat{\delta}_i^{(q+1)} = \delta_i^{(q)} - \left[ \frac{\psi(\delta_i)}{\psi'(\delta_i)} \right]^{(q)} \quad (58)$$

for  $\beta_n$  and  $\delta_i$ , respectively.

In the execution of the algorithm, each  $\beta_n$  is improved until a chosen degree of convergence is established. The same procedure is applied then to each  $\delta_i$ . Before reestimating the  $\beta_n$  with the new values of  $\delta_i$ , the constraint  $\sum_i \hat{\delta}_i = 0$  is imposed.

In a program written to execute these equations, initial estimates for all persons are set to 0. The initial estimates for the stimuli are given by  $\pm [N(I-1) - s_i]^{1/2}$ . To begin the unfolding process, some stimuli must be assigned a positive value and some a negative value, and clearly some a priori notion of the order of the stimuli is important. In practice, it would be the exception rather than the rule that no theoretical ordering would be postulated.

### Information and Standard Errors

Maximum likelihood estimation leads directly to the Fisher information function with respect to a parameter, the inverse of which is an estimate of the error variance associated with the estimate. The information is the expectation of the square of the first derivative of the likelihood function which, in general, is equal to the expectation of the negative inverse of the second derivative of the same function. In multiparameter estimation, the matrix of first and second derivatives provides the corresponding values, with the diagonals of the appropriate matrices giving the information and the error variance. This multiparameter approach is not used here because of the large size of the matrix. Therefore, as an approximation to the information and error variances, and again in accordance with practice in PIRT, the second derivatives of the univariate likelihood functions will be used.

In examining  $\phi'(\beta_n)$  and  $\psi'(\delta_i)$  it is expedient to first note that

$$\frac{\partial^2 p_{nij}}{\partial \beta_n^2} = 2(\delta_i - \delta_j)(p_{nij} - p_{nij}^2) \quad (59)$$

and that

$$\frac{\partial^2 p_{nij}}{\partial \delta_i^2} = 2(\beta_n - \delta_i)(p_{nij} - p_{nij}^2) \quad (60)$$

Considering first  $\phi'(\beta_n)$ , it is readily established from the above expressions that

$$\phi'(\beta_n) = -4 \sum_i \delta_i \sum_{j, j \neq i} (\delta_i - \delta_j)(p_{nij} - p_{nij}^2) \quad (61)$$

Further insight into the above expression can be obtained by recognizing that

$$p_{nij} - p_{nij}^2 = p_{nij}(1 - p_{nij}) = (1 - p_{nij})p_{nji} = p_{nji} - p_{nji}^2 \quad (62)$$

Then expanding Equation 61 and collecting like terms which contain the factor  $(\delta_i - \delta_j)(p_{nij} - p_{nij}^2)$  gives

$$\phi'(\beta_n) = -4 \sum_i \sum_{j,j>i} (\delta_i - \delta_j)^2 p_{nij}(1 - p_{nij}) \quad (63)$$

It is apparent from Equations 59 through 63 that the variance term  $p(1 - p)$  for a Bernoulli variable with success probability  $p$  appears as usual in such an expression. However, this term is multiplied by the square of the difference between the locations of  $\delta_i$  and  $\delta_j$ .

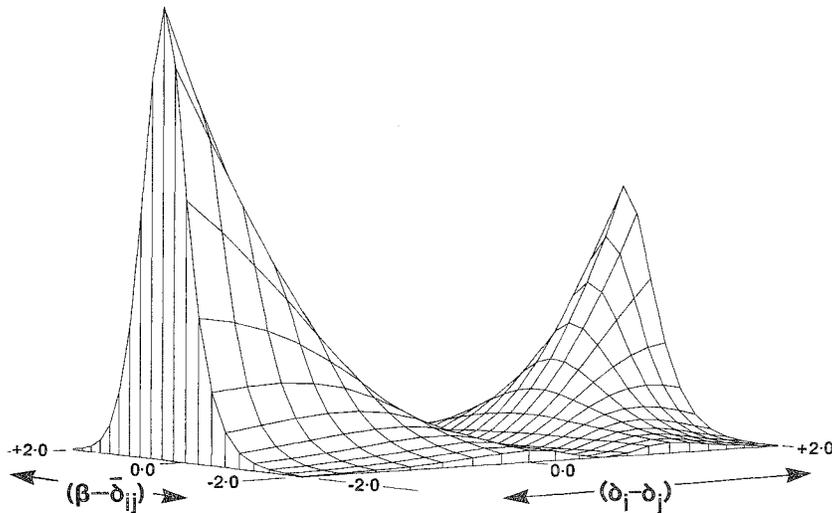
The information  $I_{\beta_n}$  for the location  $\beta_n$  is given by

$$I_{\beta_n} \equiv -\frac{\partial^2 \log L}{\partial \beta_n^2} = -\phi'(\beta_n) = 4 \sum_i \sum_{j,j>i} (\delta_i - \delta_j)^2 p_{nij}(1 - p_{nij}) \quad (64)$$

A surprising result is that the contribution to the information  $I_{\beta}$  is 0 when  $\delta_i = \delta_j$ . However, it seems intuitively clear that a choice between two stimuli which are identically located, so that each is just as likely to be selected as the other, provides no information regarding the location of the person. Apparently, some discrepancy between the locations of the stimuli is necessary so that the choice will provide information about the person's location.

To explore this issue further, consider the information from a single response of a person to two items. Denote this by  $I_{\beta|\delta_{ij}} = 4(\delta_i - \delta_j)^2 p_{nij}(1 - p_{nij})$ . Clearly,  $I_{\beta|\delta_{ij}}$  depends on both  $(\delta_i - \delta_j)$  and  $(\beta - \bar{\delta}_{ij})$ . Although  $I_{\beta|\delta_{ij}} = 0$  when  $\delta_i = \delta_j$ , it is not expected to increase monotonically as  $(\delta_i - \delta_j)$  increases. Instead, it is expected to reach a maximum at some finite value of  $(\delta_i - \delta_j)$ , decreasing when this value is exceeded. Indeed, such is the case, except when  $\beta = \bar{\delta}_{ij}$ , in which case  $I_{\beta|\delta_{ij}}$  is a quadratic function:  $(\delta_i - \delta_j)^2$ . Figure 3 shows the information surface as a function of  $(\delta_i - \delta_j)$  and  $(\beta - \bar{\delta}_{ij})$ . In this figure,  $\bar{\delta}_{ij} = 0$  without loss of generality.

**Figure 3**  
Information With Respect to  $\beta$  From the Choice Between  $i$  and  $j$



A parallel development to the above for the information  $I_{\delta_i}$  with respect to  $\delta_i$  is possible and is given by

$$I_{\delta_i} \equiv -\frac{\partial^2 \log L}{\partial \delta_i^2} = -\psi'(\delta_i) \quad , \quad (65)$$

but with a slight extra complication: Differentiating

$$\psi(\delta_i) = 2 \left[ \delta_i \left( -T_i + \sum_n \sum_{j,j \neq i} p_{nij} \right) + \sum_n \beta_n \left( s_{ni} - \sum_{j,j \neq i} p_{nij} \right) \right] = 2 \left[ -T_i \delta_i + \sum_n \beta_n s_{ni} - \sum_n \sum_{j,j \neq i} (\beta_n - \delta_i) p_{nij} \right] \quad (66)$$

gives

$$\psi'(\delta_i) = 2 \left[ -T_i - \sum_n \sum_{j,j \neq i} 2(\beta_n - \delta_i)^2 (p_{nij} - p_{nij}^2) + \sum_n \sum_{j,j \neq i} p_{nij} \right] \quad . \quad (67)$$

Unlike  $\phi'(\beta_n)$ ,  $\psi'(\delta_i)$  contains a statistic,  $-T_i$ . Therefore, the expected value  $-E[\psi'(\delta_i)]$ , the Fisher information function, is taken explicitly.

Because

$$E(T_i) = \sum_n \sum_{j,j \neq i} p_{nij} \quad , \quad (68)$$

$I_{\delta_i}$  reduces to

$$I_{\delta_i} \equiv -E[\psi'(\delta_i)] = 4 \sum_n \sum_{j,j > i} (\beta_n - \delta_i)^2 p_{nij} (1 - p_{nij}) \quad . \quad (69)$$

Equation 69 is similar to Equation 64, and it shows that the information contributed to a stimulus parameter is 0 when the person making a choice has a location value identical to that of the stimulus. At face value, this result is also surprising. It seems that the stimuli  $j$  ( $j = 1, \dots, I; j \neq i$ ) take the role of the person parameter in the information associated with the person. Therefore stimuli  $j$  ( $j = 1, \dots, I; j \neq i$ ) cannot be used to decide which is closer to any of these items, the person or item  $i$ , when the person and item  $i$  have the same location. Thus, even though the model for the probability of a choice between stimuli  $i$  and  $j$  is not symmetrical in  $i, j$ , and  $\beta$ , but only in  $i$  and  $j$ , the information associated with the estimation of  $\delta_i$  ( $i = 1, \dots, I$ ) and of  $\beta_n$  ( $n = 1, \dots, N$ ) is symmetrical.

The contribution to the relevant information is now 0 when either  $\delta_i = \delta_j$  or  $\delta_j = \beta_n$ . Again it is not expected to increase monotonically as either  $|\delta_i - \delta_j|$  or  $|\beta_n - \delta_j|$  increases, but to reach a maximum at finite values of  $|\delta_i - \delta_j|$  and  $|\beta_n - \delta_j|$ .

Let the information for the estimate of  $\delta_i$  be denoted by

$$I_{\delta_i|\beta_n, \delta_j} = 4(\beta_n - \delta_i)^2 p_{nij} (1 - p_{nij}) \quad . \quad (70)$$

Figure 4 shows this information surface as a function of  $|\delta_i - \delta_j|$  and  $(\beta_n - \bar{\delta}_j)$ .

Finally, the variances of the estimates of  $\beta_n$  and  $\delta_i$  based on Equations 64 and 69 are estimated by

$$\hat{\sigma}_{\beta_n}^2 = \left[ 4 \sum_i \sum_{j,j > i} (\delta_i - \delta_j)^2 p_{nij} (1 - p_{nij}) \right]^{-1} \quad (71)$$

and

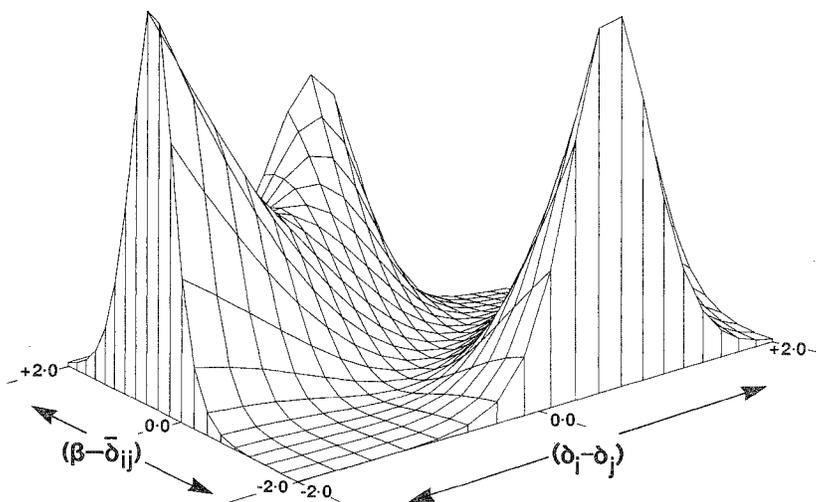
$$\hat{\sigma}_{\delta_i}^2 = \left[ 4 \sum_i \sum_{j,j > i} (\beta_n - \delta_i)^2 p_{nij} (1 - p_{nij}) \right]^{-1} \quad , \quad (72)$$

respectively. Examples of these values are shown in the simulation and real data studies.

### A Simulation Study

The model and the solution algorithm described in the previous section indicates that estimating

**Figure 4**  
Information With Respect to  $\delta_i$  From the Choice Between  $i$  and  $j$



parameters should be quite feasible, although it does not exhaustively deal with all issues that are important in models with both incidental and structural parameters (Andersen, 1973).

The simulation involved seven stimuli and 250 persons who made comparisons with respect to each pair of stimuli according to the model of Equations 16, 20, and 23. The seven stimuli were symmetrically centered, though not uniformly spaced, about a scale value of 0 with a range from  $-2$  to  $+2$ . The location values of the persons were distributed normally with mean 0 and standard deviation 1. A convergence criterion of .01 was used in each of five replications of the simulation. Seven is a relatively large number of stimuli for a complete pairwise preference design, as it involves 21 comparisons for each person.

#### Recovery of Stimulus Parameters

Table 5 shows the generating parameters of the items, and the estimated parameters and their standard errors in each replication. The estimates are clearly of the correct order of magnitude. The estimates tend to be consistently spread more than the generating parameters, and by more than would be expected just because the estimates contain an error component. However, this result is not surprising. The model has both person and stimulus parameters similar to that of the SLM in which the direct MLE procedure provides inconsistent estimates (Andersen, 1973). The nature of the inconsistency in the SLM is such that the stimulus parameter estimates have a systematically greater spread than the parameters. A correction factor for these estimates is available with the SLM (Wright & Douglas, 1977) but this issue is not explored with the model presented here.

#### Recovery of Person Parameters

Table 6 shows the quality of the recovery of the person parameters as summarized by the following statistics:  $r(\underline{\beta}, \hat{\underline{\beta}})$ , the correlation of estimated and generating parameters;  $\bar{\beta}$ , the mean of generating parameters;  $\hat{\bar{\beta}}$ , the mean of estimated parameters;  $\sigma_{\beta}^2$ , the variance of generating parameters;  $\sigma_{\hat{\beta}}^2$ , the

Table 5  
 Generated ( $\delta$ ) and Estimated ( $\hat{\delta}$ ) Stimulus Parameters  
 and Their Standard Errors (SE) in Five Replications

Stimulus Number	$\delta$		$\hat{\delta}$				
			1	2	3	4	5
1	-2.0	Est	-2.164	-2.167	-2.225	-2.387	-2.369
		SE	.035	.038	.039	.037	.039
2	-1.0	Est	-1.011	-1.145	-1.142	-1.213	-.981
		SE	.038	.037	.041	.039	.040
3	-.5	Est	-.572	-.643	-.386	-.605	-.417
		SE	.044	.042	.044	.047	.042
4	0.0	Est	.030	.032	.054	.231	.009
		SE	.052	.048	.044	.047	.045
5	.5	Est	.568	.624	.536	.658	.516
		SE	.042	.041	.040	.039	.042
6	1.0	Est	1.071	1.130	1.038	1.090	.991
		SE	.035	.037	.038	.035	.039
7	2.0	Est	2.078	2.169	2.125	2.227	2.250
		SE	.036	.038	.039	.039	.038
Mean	0.0		0.000	0.000	0.000	0.000	0.000
Variance	1.5		1.689	1.827	1.755	2.024	1.865
Correlation			1.000	.999	.998	.997	.997
Regression of $\delta$ on $\hat{\delta}$			.940	.910	.920	.860	.890

variance of estimated parameters;  $\bar{\sigma}_\beta^2$ , the average error variance within each person parameter estimate;  $\hat{\sigma}_\beta^2 = \hat{\sigma}_\beta^2 - \bar{\sigma}_{\beta n}^2$ , the estimate of true variance among person parameters; and  $\hat{\sigma}_\beta^2 - \sigma_\beta^2$ , the difference between the variance of the generating parameters and the estimate of the true variance of the generating parameters.

The correlations between the generating and estimated parameters are extremely high, given the errors of the estimates. Note that  $\sigma_\beta^2$  is smaller than the value set in the simulation, namely 1. This reduction can be attributed to the elimination of about 20 persons from each set; these persons were eliminated because they had extreme observed response patterns which prevented obtaining a finite estimate of their locations. Extreme response patterns are those in which a person responds in a perfect unfolding pattern which corresponds to a location outside the range of the stimuli. For example, the response patterns of Persons 1 and 6 associated with Figure 1 and shown in Table 4 correspond to extreme patterns. Such extreme patterns are most likely to arise with persons with extreme locations, thus reducing the variance of the remaining person locations.

The variances of the estimates  $\hat{\sigma}_\beta^2$  are clearly greater than those of the generating parameters  $\sigma_\beta^2$ . The increases result in part from the errors associated with the estimates. The differences between these two quantities show that except for one simulation, these values still exceed the variance of the generating parameters, but only marginally. Thus any inconsistencies in the estimates of the person parameters are likely to be minimal. Also, if more than seven stimuli were used the effect would be even smaller.

### Example Applications

#### Example 1: Students' Expectations of Grades

Students in an introductory course on educational measurement and statistics were asked to make comparisons between each pair of possible grades, and to indicate the grade in each pair that they were most likely to achieve. The possible grades were A+, A, B+, B, C+, C, N+.

Table 6  
Relationships Between Generated and Estimated  
Person Parameters in Five Simulation Studies

Statistic	Simulation				
	1	2	3	4	5
$r(\beta, \hat{\beta})$	.922	.924	.937	.907	.921
$\bar{\beta}$	-.052	-.020	.011	-.015	-.009
$\hat{\beta}$	-.067	-.019	.025	-.029	-.023
$\sigma_{\beta}^2$	.690	.776	.867	.808	.755
$\sigma_{\hat{\beta}}^2$	.734	.989	1.099	.984	.990
$\bar{\sigma}_{\hat{\beta}}^2$	.067	.080	.090	.077	.074
$\hat{\sigma}_{\beta}^2 = \hat{\sigma}_{\beta}^2 - \bar{\sigma}_{\hat{\beta}}^2$	.667	.909	1.009	.907	.916
$\hat{\sigma}_{\beta}^2 - \sigma_{\beta}^2$	-.023	.133	.142	.099	.161
Number Deleted	19	18	21	23	22

Table 7 shows the frequency of choices among pairs of grades, and the expected frequencies according to the model and the estimated parameters. Table 7 also shows the scale estimate of each of the grades, based on the data from 66 students remaining after four were deleted. It is both interesting and reassuring that the pass grades, those between C and A+, show a symmetry with the greatest distance perceived to be between B+ and B.

As a guide to the statistical quality of the fit between the model and the data, a Pearson  $\chi^2$  based on the difference between the observed and expected frequencies in the cells can be computed as

Table 7  
Pair Comparisons of Grade Expectations: Variances of  
Expected Frequencies Above the Diagonal, Observed (O)  
and Expected (E) Frequencies Below the Diagonal

Grade	Grade							Scale Value
	A+	A	B+	B	C+	C	N+	
A+		2.04	.46	1.28	2.13	2.73	4.69	3.10
A	O 64		2.95	3.28	4.97	6.29	1.63	2.36
	E 63.81							
B+	O 65	63		9.85	6.25	2.61	1.42	1.21
	E 65.47	62.33						
B	O 62	56	35		6.74	2.82	.73	-.46
	E 62.72	57.06	39.45					
C+	O 53	45	23	12		7.81	1.01	-1.41
	E 55.77	45.89	21.49	12.27				
C	O 52	35	15	6	5		1.58	-1.93
	E 50.26	33.83	14.15	6.76	10.73			
N+	O 28	14	8	5	1	0		-2.87
	E 27.90	14.32	8.80	3.15	2.01	2.19		

$$\chi^2 = \sum_{i,j>j} \frac{\left(f_{ij} - \sum_n p_{nij}\right)^2}{\sum_n p_{nij}(1 - p_{nij})} \tag{73}$$

with  $(I - 1)(I - 2)$  degrees of freedom. The resulting  $\chi^2$  of 22.79 with 15  $df$  ( $p > .09$ ) indicates that clearly, according to this criterion, the data accord quite well with the model.

Figure 5 shows the estimated locations of the grades and a frequency distribution of the locations of the persons. It seems that most students see themselves as attaining a grade between B and B+.

**Example 2: Attitudes Toward Capital Punishment**

The second example concerns attitude to capital punishment. The statements used for the measurement of this attitude were a subset of those scaled according to Thurstone's pair-comparison method and subsequently according to a method proposed by Mostellor (Wohlwill, 1963). In the latter case, the scaling was done by considering the persons' direct Agree or Disagree responses to statements.

Students ( $N = 49$ ) in the same class as that of Example 1 were asked to indicate which of each pair of statements reflected their opinion on capital punishment more closely, where every statement was paired with every other statement.

Table 8 shows the statements and the scale values and standard errors obtained according to Equations 16, 20, and 23. In addition, scale values for statements obtained by the Thurstone method but scaled to have the same mean (0.0) and standard deviation (1.10) as those from Equations 16, 20, and 23 are shown. Table 9 shows the frequency of each choice, and the expected frequencies on the basis of the model, with  $\chi^2$  of 14.67 ( $df = 6, p > .02$ ) based on  $N = 46$ . From Table 8, it is clear that the scale values from Thurstone's data and method of scaling are similar to the scale values of the data obtained from the data described here and analyzed according to Equation 6. From Table 9, it is also evident that the data are of the unfolding type.

Finally, Figure 6 shows the distribution of attitude toward capital punishment in the group who responded. Perhaps not surprisingly, attitudes cover a wide range, from very much for to very much against capital punishment.

**Figure 5**  
 Scale Values of Grades and the Distribution of Expected Grades

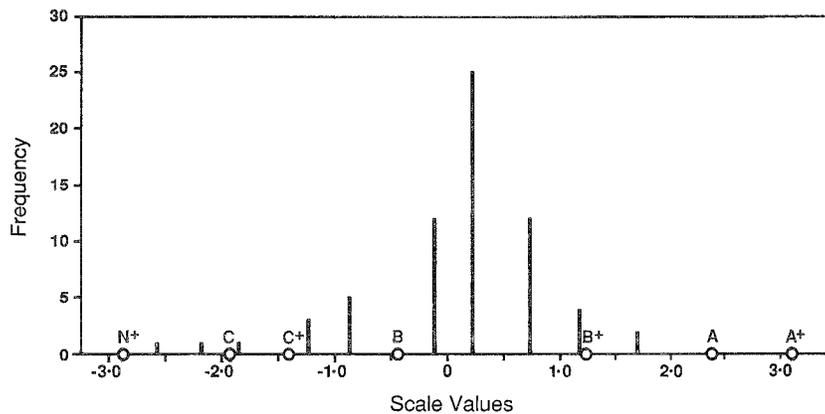


Table 8  
Statements Expressing an Attitude Toward Capital Punishment,  
Scale Value Estimates and Their Standard Errors (SE)

Statement	Scale Values		SE
	Thurstone's Method	Preference Method	
1. Capital punishment is one of the most hideous practices of our time.	-1.58	-1.52	.08
2. I don't believe in capital punishment, but I am not sure it isn't necessary.	.11	-.01	.17
3. Capital punishment is not an effective deterrent to crime.	-.80	-.88	.10
4. Until we find a more civilized way to prevent crime, we must have capital punishment.	.75	1.15	.07
5. Capital punishment gives the criminal what he deserves.	1.52	1.27	.07

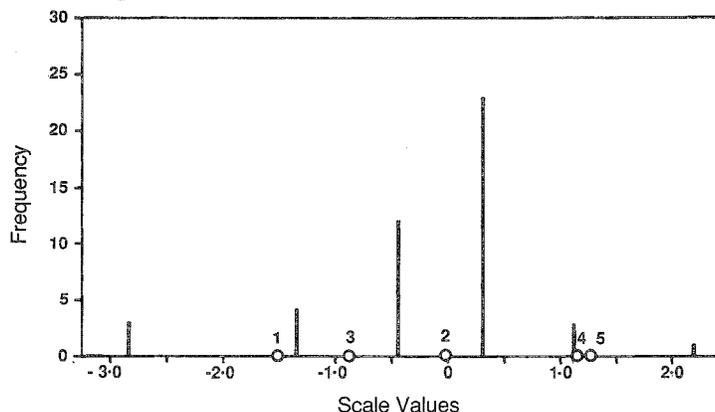
It is rather encouraging for the stability of the scale to note that the two sets of scale values of the statements are so similar. It is particularly encouraging when the differences in the data collection and analysis are noted:

1. The data were collected almost 50 years apart;
2. They were collected in two different countries and in different types of cities;
3. Thurstone's method assumed the cumulative mechanism operated, in which statements were compared for affective values irrespective of the person's attitude, while the method described here assumed that the unfolding method operated, in which the person's attitude governed the choice; and
4. Correspondingly different but specific models were used successfully to estimate the scale values.

Table 9  
Pair Preferences Among Statements Expressing an Attitude  
Toward Capital Punishment: Variances of Expected  
Frequencies Above the Diagonal, Observed (O) and  
Expected (E) Frequencies Below the Diagonal

State- ment	Statement					Scale Value
	1	2	3	4	5	
1		3.76	6.39	3.86	4.00	-1.52
2	O 37		7.54	6.54	5.90	-.01
	E 34.92					
3	O 31	20		5.80	5.61	-.87
	E 34.56	17.59				
4	O 30	12	16		11.07	1.14
	E 27.81	13.02	20.95			
5	O 29	14	17	13		1.26
	E 26.68	11.74	18.68	19.55		

**Figure 6**  
 Scale Values of Statements Concerning  
 Capital Punishment and the Distribution of Attitudes



### Discussion

The model presented here makes three modifications to Thurstone's discriminial process: (1) Persons are identified explicitly, rather than selected randomly from some population; (2) The location of the stimulus is assumed to be governed by the square of the difference between the person and the stimulus locations rather than only by the location of the stimulus; and (3) the discriminial process is assumed to be distributed as the double exponential, which then uses a logistic choice function rather than a cumulative normal.

Concentrating on a single dimension and deriving the model from first principles using two distinct approaches clarified properties of the unfolding choice mechanism, especially in relation to the alternative cumulative choice mechanism. In the cumulative mechanism it is assumed that the stimuli are chosen only according to their perceived relative locations on the particular dimension, whereas in the unfolding mechanism it is assumed that the stimuli are chosen according to their perceived distances relative to the location of the person making the choice.

The inability to eliminate the location of the person in this model is consistent with intuitive notions of unfolding, in which the location of the person relative to the midpoint between the locations of the two stimuli plays a central role. Like most other models, this model incorporates the distance between the person and the midpoint of the two stimuli. However, unlike some other models, this model also incorporates the distance between the two stimuli. If a person is not located at the midpoint between the stimuli, then the greater the distance between two stimuli the greater the probability that one will be chosen over the other for any particular location of the person.

The model incorporates the two distances very simply: The probability of the choice of stimulus  $i$  over stimulus  $j$  is governed by the product of these directed distances. The model does not have these two factors postulated in a simple structure but it is derived from prior plausible assumptions regarding the process induced when a person encounters a stimulus.

One purpose for the construction of models, which inevitably do not account for all aspects of the phenomena under study, is that they add insight into certain features rendered explicit by the chosen model. The maximum likelihood estimates and the associated information functions do provide further

insight into the unfolding choice mechanism. According to the model, if the two stimuli have identical locations, then not only are the stimuli equally likely to be chosen but this choice provides no information about the location of the person. Analogously, if a person is located at the same position as one of the stimuli, then a choice between that stimulus and any other provides no information about the stimulus which has the identical location to the person.

The first result appeals intuitively in that if two stimuli have identical locations, then the probabilities of the choices of the stimuli are equal, irrespective of the location of the person. The second result, though analogous, is not as easy to appreciate. However, if one stimulus has an identical location to that of the person, then no second stimulus can be used to distinguish between the person and the first stimulus. That is, no information can be obtained about the first stimulus. These conclusions seem partly intuitively consistent with the model, but also raise new issues in the unfolding response mechanism.

Whether an unfolding or a cumulative choice mechanism is operating is an empirical matter, and the nature of the task should indicate which is the most likely. Because of the simple operationalization of the model, it should be a relatively straightforward matter to test whether an unfolding or cumulative mechanism is operating primarily in any set of real data. This can be done by applying both the model studied in the paper and the Bradley-Terry-Luce model and comparing the results.

### References

- Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31–44.
- Andrich, D. (1978a). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, 2, 451–462.
- Andrich, D. (1978b). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology*, 31, 84–98.
- Andrich, D. (1985a). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological methodology* (pp. 33–80). San Francisco: Jossey-Bass.
- Andrich, D. (1985b). The construction of a probabilistic model for the psychological scaling of unfolding choice data. In E. E. Roskam (Ed.), *Measurement and personality assessment*. Amsterdam: Elsevier.
- Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement*, 12, 33–51.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51.
- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs I: The method of paired comparisons. *Biometrika*, 39, 324–45.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the social sciences* (Vol. 1, pp. 105–153). New York: Seminar Press.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145–158.
- Coombs, C. H. (1964). Some symmetries and dualities among measurement data matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 42–59). Toronto: Holt, Rinehart and Winston.
- Coombs, C. H., & Avrunin, C. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review*, 84, 216–230.
- Coombs, C. H., & Smith, J. E. K. (1973). On the detection of structure in attitudes and developmental process. *Psychological Review*, 5, 337–351.
- Cooper, L. G., & Nakanishi, M. (1983). Two logit models for external preferences. *Psychometrika*, 48, 607–620.
- Davison, M. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika*, 42, 523–548.
- DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement*, 10, 247–264.
- Jansen, P. G. W. (1984). Relationships between the Thurstone, Coombs, and Rasch approaches to item scaling. *Applied Psychological Measurement*, 8, 373–383.

- Johnson, N. L., & Kotz, S. (1972). *Distributions in statistics: Continuous univariate distributions*, 3. New York: Wiley.
- Jensema, C. J. (1974). An application of latent trait mental test theory. *British Journal of Mathematical and Statistical Psychology*, 27, 29–48.
- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three parameter logistic model. *Educational and Psychological Measurement*, 28, 989–1200.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R. D. (1977). Thurstone's discriminial process fifty years later. *Psychometrika*, 42, 461–489.
- Poole, K. T. (1984). Least squares metric, unidimensional unfolding. *Psychometrika*, 49, 311–323.
- Ramsay, J. O. (1980). The joint analysis of direct response ratings, pairwise preferences and dissimilarities. *Psychometrika*, 45, 149–165.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press, 1980.]
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321–334). Berkeley CA: University of California Press.
- Schönemann, P. H. (1970). On metric multidimensional unfolding. *Psychometrika*, 35, 349–366.
- Schönemann, P. H., & Wang, M. M. (1972). An individual difference model for the multidimensional analysis of preference data. *Psychometrika*, 37, 275–309.
- Sixtl, F. (1973). Probabilistic unfolding. *Psychometrika*, 38, 235–248.
- Tanzer, N. (1983). *Messung von Einstellungen mittels probabilistischer Meßmodelle*. Unpublished doctoral dissertation, University of Vienna.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 278–286.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–54.
- Thurstone, L. L. (1931). Rank order as a psychophysical method. *Journal of Experimental Psychology*, 14, 182–201.
- Thurstone, L. L. (1959). *The measurement of values*. Chicago: University of Chicago Press.
- Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude*. Chicago: University of Chicago Press.
- Wohlwill, J. F. (1963). The measurement of scalability for non-cumulative items. *Educational and Psychological Measurement*, 23, 543–55.
- Wright, B. D., & Douglas, G. A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement*, 37, 47–60.
- Yellott, J. I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15, 109–144.

#### Acknowledgments

Annette Mercer worked through the equations in this paper, Alan Lyne wrote the corresponding FORTRAN program, and John Horgan produced the three-dimensional graphs.

#### Author's Address

Send requests for reprints or further information to David Andrich, School of Education, Murdoch University, Murdoch, Western Australia 6150, Australia.