MATHÉMATIQUES ET SCIENCES HUMAINES

G. DE SOETE

On the relation between two generalized cases of Thurstone's law of comparative judgment

Mathématiques et sciences humaines, tome 81 (1983), p. 47-57.
<http://www.numdam.org/item?id=MSH_1983__81__47_0>

Math. Sci. hum. (21e année, n°81, *1983*, p.47-57)

ON THE RELATION BETWEEN TWO GENERALIZED CASES OF THURSTONE'S LAW OF COMPARATIVE JUDGMENT

G. DE SOETE*

Choice behavior is of central interest to many social scientists. The observation that choice behavior is often inconsistent has led psychologists to develop probabilistic models to represent pairwise choice data. These models allow one to derive from paired comparison data numerical estimates of the utilities of the choice objects involved. It should be noticed that these psychological models may be very useful in many other social science disciplines, ranging from marketing to political science.

I. THURSTONE'S LAW OF COMPARATIVE JUDGMENT

One of the most popular probabilistic models for representing pairwise choice data is Thurstone's (1927) Law of Comparative Judgment (LCJ). According to Thurstone, the utility of a choice object is not constant but fluctuates from time to time. Therefore, in his model the utility of a stimulus i (i = 1, n) is represented as a random variable U_i. When presented a pair of stimuli, the subject is supposed to choose the stimulus with the highest momentary utility. Consequently, the probability of preferring

*"Aspirant" of the Belgian "Nationaal Fonds voor Wetenschappelijk Onderzoek" at the Department of Psychology, University of Ghent, Belgium.
The author is indebted to Yoshio Takane for his critical comments on an earlier draft of this paper.

i to j ($i \neq j$), written as p_{ij}, is

$$p_{ij} = \text{Prob}\ (U_i > U_j). \qquad (1)$$

Assuming that $\underline{U}' = (U_1, \ldots, U_n)$ has a multivariate normal distribution[1]

$$\underline{U} \sim N(\underline{\mu}, \Sigma),$$

equation 1 becomes

$$p_{ij} = \Phi\left(\frac{\mu_i - \mu_j}{\delta_{ij}}\right) \qquad (2)$$

where Φ denotes the standard normal distribution function and δ_{ij} the so-called comparatal dispersion (Gulliksen, 1958), i.e.,

$$\delta^2_{ij} = \text{Var}\ (U_i - U_j)$$
$$= \underline{e}_{(ij)}{}'\Sigma\underline{e}_{(ij)}$$

where $\underline{e}_{(ij)}$ is a n-element column vector whose k'th component is defined by $\delta^{ik} - \delta^{jk}$ (with $\delta^{\cdot\cdot}$ Kronecker delta).

Equation 2 constitutes Thurstone's complete LCJ. The model incorporates $n + n(n+1)/2$ parameters. However, any transformation of the form

$$\underline{\mu} \rightarrow a\underline{\mu} + b\underline{1}$$
$$\Sigma \rightarrow a^2\Sigma + \underline{c}\underline{1}' + \underline{1}\underline{c}'$$

where a is a positive scalar, $\underline{c}$ an arbitrary n-component vector, and $\underline{1}$ a vector of n ones, leaves the choice probabilities invariant. Therefore, the effective number of parameters (or the degrees of freedom for the model) reduces to $n(n+1)/2 - 2$. Since this number still exceeds the number of observed choice proportions $[n(n-1)/2]$, restrictions need to be imposed on the parameters in order for the model to be of any practical use. The most frequently used restriction in empirical research requires δ_{ij} to be constant for all (i,j). This is for instance the case when Σ is of the form

$$\Sigma = aI + \underline{c}\underline{1}' + \underline{1}\underline{c}'$$

where a is a positive scalar, $\underline{c}$ an arbitrary n-component vector, and I an

[1]Strictly speaking, all that is needed to arrive at equation 2 is marginal univariate normality of $\underline{U}$. However, assuming multivariate normality is theoretically more attractive.

identity matrix of order n. When this constraint is imposed on (2), the model is referred to as the LCJ Case V. Without loss of generality, Case V can be written as

$$p_{ij} = \Phi(\mu_i - \mu_j). \tag{3}$$

The effective number of parameters in this model amounts to n - 1.

Equation 3 states that the choice probabilities are a strictly increasing function of the difference between the mean utilities of the two alternatives. Experimental research (Becker, DeGroot, and Marschak, 1963; Krantz, 1967; Rumelhart and Greeno, 1971; Tversky and Russo, 1969; Tversky and Sattath, 1979) suggests, however, that empirical choice proportions are influenced not only by differences in utility, but also - to some extent - by the similarity or comparability of the stimuli. Subjects tend to be rather indifferent between very dissimilar stimuli, even if the stimuli differ a lot in utility. Similar stimuli, on the other hand, tend to give rise to extreme choice proportions, even when they do not differ that much in utility. In a series of studies conducted by Sjöberg (1975, 1977, 1980; Sjöberg and Capozza, 1975), it was found that the discriminal dispersion in (2) is related to the dissimilarity between the alternatives. Consequently, constraining the discriminal dispersions to be equal to each other - as in Case V - is from a psychological point of view not very realistic. Therefore, researchers became interested in developing Thurstonian models which *do* allow the comparatal dispersions to reflect the dissimilarity between the stimuli. Recently, two such models have been proposed. The first one is called by Takane (1980) the *factorial model of comparative judgment*, while the second one is known as the *wandering vector model* (Carroll, 1980). Because these cases of the LCJ differ from the five cases discussed by Thurstone (1927), they are called generalized cases. These generalized cases attempt to extract from a single paired comparisons matrix information not only about the mean utilities of the alternatives, but also about the similarities between the choice objects. In the next sections we discuss the two models as well as how they relate to each other. In the final section, both models are applied for comparative purposes on a data set obtained by Rumelhart and Greeno (1971).

II. THE FACTORIAL MODEL OF COMPARATIVE JUDGMENT

The factorial model of comparative judgment has been proposed by Takane (1980) and by Heiser and de Leeuw (1981). In the factorial model, the covariance matrix Σ is required to have a prescribed rank r ($r < n$). Since Σ is positive semi-definite, it can be decomposed as follows

$$\Sigma = XX' \tag{4}$$

where X is a n x r matrix. Equation 4 is a special case of a general model for constraining Σ discussed by Arbuckle and Nugent (1973, eq. 23). Given this restriction, the comparative variance δ^2_{ij} becomes

$$\begin{aligned}\delta^2_{ij} &= \underline{e}_{(ij)}'XX'\underline{e}_{(ij)} \\ &= (\underline{x}_i - \underline{x}_j)'(\underline{x}_i - \underline{x}_j) \\ &= d^2_{ij}(X)\end{aligned}$$

where the column vector $\underline{x}_i$ stands for the i'th row of X. Consequently, the comparatal dispersions may be interpreted as Euclidean distances in a r-dimensional space. In this space each choice object i is represented as a point $\underline{x}_i$. This space can be thought of as a geometric representation of the subject's cognitive map which presumably influenced his or her preference judgments.

The parameters to be estimated in the factorial model are $\underline{\mu}$ and X. The choice probabilities p_{ij} are invariant under any transformation of the form

$$\begin{aligned}\underline{\mu} &\rightarrow a\underline{\mu} + b\underline{1} \\ X &\rightarrow aXT + \underline{1}\underline{c}'\end{aligned}$$

where a is a positive scalar, T any orthogonal matrix of order r, and $\underline{c}$ an arbitrary r-component vector. The effective number of parameters in the factorial model is $n + nr - r(r+1)/2 - 2$ (and not $nr - r(r+1)/2 - 1$ as Takane (1980, p. 194) incorrectly states[2]).

An algorithm for obtaining maximum likelihood estimates of $\underline{\mu}$ and X has been presented by Takane (1980), while Heiser and de Leeuw (1981) discuss how to fit the factorial model by least squares methods using the SMACOF approach (cf. de Leeuw and Heiser, 1980).

III. THE WANDERING VECTOR MODEL

The wandering vector model, originally introduced by Carroll (1980), attempts a geometric representation of both the subject and the choice objects in a joint r-dimensional space. Each object i is represented as a point $\underline{x}_i$, while the subject is represented as a vector originating from the origin. According to the model, each time the subject is presented a pair of stimuli he or she chooses the stimulus with the (algebraically) largest orthogonal projection on the subject vector. The wandering vector model does not assume that the position of the subject vector is constant, but rather that it changes ("wanders") somewhat from time to time. Formally, it is assumed that $\underline{y}$, the terminus of the subject vector, follows a multivariate normal distribution. Without loss of generality the variance-covariance matrix of this distribution can be set equal to the identity matrix, i.e.,

$$\underline{y} \sim N(\underline{v}, I). \tag{5}$$

When presented a pair of stimuli (i,j), the subject is supposed to sample a $\underline{y}$ from this distribution. The subject then chooses i whenever the orthogonal projection of $\underline{x}_i$ on $\underline{y}$ exceeds the orthogonal projection of $\underline{x}_j$ on $\underline{y}$, or whenever

$$\underline{x}_i'\underline{y} > \underline{x}_j'\underline{y}.$$

[2]The degrees of freedom given in Table 2 of Takane (1980) are however correct.

Consequently, the probability of preferring i to j is

$$p_{ij} = \text{Prob}\ [(\underline{x}_i - \underline{x}_j)'\underline{y} > 0]. \qquad (6)$$

Since it follows from (5) that $(\underline{x}_i - \underline{x}_j)'\underline{y}$ is normally distributed with mean $(\underline{x}_i - \underline{x}_j)'\underline{v}$ and variance $d^2_{ij}(X)$, equation (6) becomes

$$p_{ij} = \Phi\ \left(\frac{(\underline{x}_i - \underline{x}_j)'\underline{v}}{d_{ij}(X)} \right). \qquad (7)$$

These choice probabilities are not affected by any transformation of X of the form

$$X \rightarrow aXT + \underline{1}\underline{c}'$$

where a is a positive scalar, T any orthogonal matrix of order r, and $\underline{c}$ an arbitrary r-component vector. Consequently, the number of degrees of freedom for the wandering vector model is $nr - r(r-1)/2 - 1$. For several interesting extensions of the model as well as for a maximum likelihood estimation procedure, the reader is referred to De Soete and Carroll (1982).

IV. THE RELATION BETWEEN THE TWO MODELS

Heiser and de Leeuw (1981, p. 50) assert that the factorial model of comparative judgment and the wandering vector model are essentially equivalent. This is, however, not completely true. From the description of both models, it should be clear that the wandering vector model is a special case of the factorial model. More specifically, if we constrain the mean utilities $\underline{\mu}$ in the factorial model to be a linear function of the stimulus coordinates X, we obtain the wandering vector model. Or in other words, the wandering vector model is only equivalent to the factorial model if we impose on the latter the constraint

$$\underline{\mu} = X\underline{v}.$$

If this constraint is consistent with the data, the wandering vector model should be preferred to the factorial model since it involves $n - r - 1$ less parameters. An important advantage of the wandering vector model is that the stimulus configuration conveys information about the mean utilities of the choice objects.

V. ILLUSTRATIVE APPLICATION

Rumelhart and Greeno (1971) obtained preference judgments from 234 undergraduates about nine celebrities. These celebrities consisted of three politicians (L.B. Johnson (LJ), Harold Wilson (HW), Charles De Gaulle (CD)), three athletes (Johnny Unitas (JU), Carl Yastrzemski (CY), A.J. Foyt (AF)), and three movie stars (Brigitte Bardot (BB), Elizabeth Taylor (ET), Sophia Loren (SL)). The subjects were treated as replications of each other. Takane (1980) fitted the LCJ Case V as well as the two-dimensional factorial model of comparative judgment to these data using maximum likelihood estimation. In order to compare the two generalized cases of the LCJ with each other, the same data were analyzed according to the wandering vector model in two dimensions using the maximum likelihood estimation procedure described by De Soete and Carroll (1982).

Table 1. Summary of the analyses on the Rumelhart and Greeno (1971) data

model	ln L (+5000)	effective no. of parameters	AIC (-10000)	Test against null model χ^2	d.f.
null model	-310.6	36	693.3		
LCJ Case V	-351.4[a]	8	718.7	81.6*	28
factorial model (r = 2)	-314.1[a]	22	672.1	7.0	14
wandering vector model (r = 2)	-318.4	16	668.8	15.4	20

[a]Obtained by Takane. These values are not exactly identical to those reported in Table 2 of Takane (1980) due to an error in the data on which the latter table is based.

*$p < 0.001$

Table 1 summarizes the results of the analyses. The null model refers to a completely saturated model in which no structural constraints are imposed on the p_{ij}. The three other models are special cases of this null model.

Whenever two models are hierarchically related to each other, it is possible to devise a likelihood ratio test. More specifically, suppose that model ω is subsumed under model Ω and that the maximum of the likelihood function obtained under both models is written respectively as $\hat{L}_\omega$ and $\hat{L}_\Omega$, then under the null hypothesis that ω fits the data equally well as Ω, the statistic

$$- 2 \ln (\hat{L}_\omega/\hat{L}_\Omega)$$

is asymptotically distributed as a chi-square with $m_\Omega - m_\omega$ degrees of freedom, where m_Ω and m_ω denote the degrees of freedom for respectively model Ω and ω. Another statistic that is useful for comparing the goodness-of-fit of different models, is Akaike's (1977) AIC which is defined for model Ω as

$$AIC_\Omega = - 2 \ln \hat{L}_\Omega + 2 m_\Omega.$$

This statistic has the advantage of explicitly correcting for the gain in goodness-of-fit due to an increased number of parameters. The smaller the AIC, the better the fit between the data and the model.

As can be inferred from the tests against the null model given in Table 1, both the factorial model and the wandering vector model give a good account of the data, while the LCJ Case V should be rejected. Since the wandering vector model is subsumed under the factorial model, the two can be compared with each other by means of a likelihood ratio test. The relevant chi-square statistic has 6 degrees of freedom and amounts to 8.6 which is not significant. This suggests that the wandering vector model fits the data equally well as the factorial model which has however six parameters more. Consequently, we can conclude that (of the models considered here) the wandering vector model gives the most parsimonious representation of the Rumelhart and Greeno (1971) data. The fact that the smallest AIC is obtained for the wandering vector model further confirms this conclusion.

A plot of the configuration obtained with the wandering vector model is presented in Figure 1. In this plot the arrow points in the direction of the subject vector. The three stimulus clusters can clearly be distinguished.

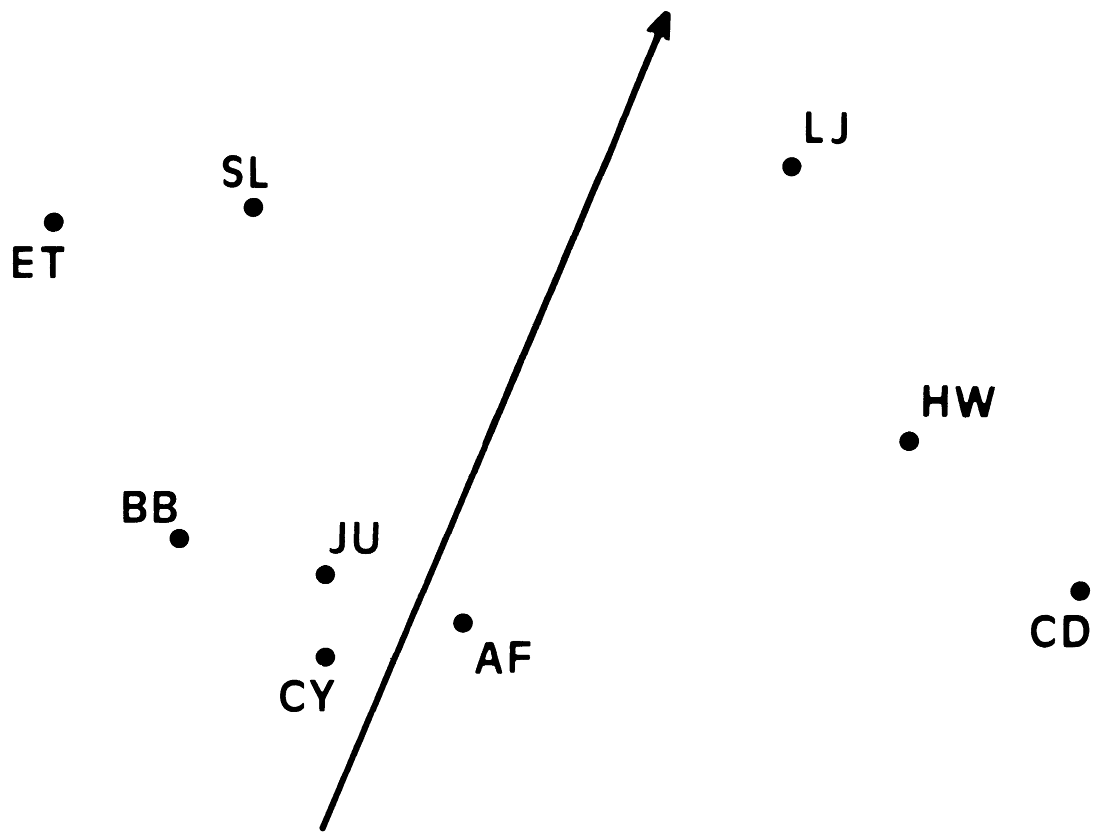

Figure 1. Configuration obtained for the Rumelhart and Greeno (1971) data with the wandering vector model in two dimensions

VI. RECAPITULATION

In this paper we discussed the relation between two generalized cases of Thurstone's Law of Comparative Judgment, namely the factorial model of comparative judgment proposed by Takane (1980) and Heiser and de Leeuw (1981) and the wandering vector model suggested by Carroll (1980). It was shown that the wandering vector model is a special case of the factorial model. Furthermore we argued that whenever the two models fit the data equally well (as is the case with the Rumelhart and Greeno data) the wandering vector model should be preferred because it gives a more parsimonious account of the data.

BIBLIOGRAPHY

(1) AKAIKE H., "On entropy maximization principle", In: Krishnaiah P.R. (ed.), Applications of statistics, Amsterdam, North-Holland Publishing Company, 1977.

(2) ARBUCKLE J. and NUGENT J.H., "A general procedure for parameter estimation for the law of comparative judgment", Brit. J. of Math. and Stat. Psych., 26 (1973), 240-260.

(3) BECKER G.M., DEGROOT M.H. and MARSCHAK J., "Probabilities of choice among very similar objects", Behav. Sci., 8 (1963), 306-311.

(4) CARROLL J.D., "Models and methods for multidimensional analysis of preferential choice (or other dominance) data", In: Lantermann E.D. and Feger H. (eds.), Similarity and choice, Bern, Huber Verlag, 1980.

(5) DE LEEUW J. and HEISER W., "Multidimensional scaling with restrictions on the configuration", In: Krishnaiah P.R. (ed.), Multivariate analysis V, Amsterdam, North-Holland Publishing Company, 1980.

(6) DE SOETE G. and CARROLL J.D., "A maximum likelihood estimation method for fitting the wandering vector model", Paper presented at the annual meeting of the Psychometric Society, Montréal, Canada, 1982.

(7) GULLIKSEN H., "Comparatal dispersion, a measure of accuracy of judg-

ment", Psychometrika, 23 (1958), 137-150.

(8) HEISER W.J. and DE LEEUW, J., "Multidimensional mapping of preference data", Math. Sci. hum., 19 (1981), 39-96.

(9) KRANTZ D.H., "Rational distance function for multidimensional scaling" J. of Math. Psych., 4 (1967), 226-245.

(10) RUMELHART D.L. and GREENO J.G., "Similarity between stimuli: an experimental test of the Luce and Restle choice models", J. of Math. Psych., 8 (1971), 370-381.

(11) SJOBERG L., "Uncertainty of comparative judgment and multidimensional structure", Multiv. Behav. Res., 11 (1975), 207-218.

(12) SJOBERG L., "Choice frequency and similarity", Scandinavian J. of Psych., 18 (1977), 103-115.

(13) SJOBERG L., "Similarity and correlation", In: Lantermann E.D. and Feger H. (eds.), Similarity and choice, Bern, Huber Verlag, 1980.

(14) SJOBERG L. and CAPOZZA D., "Preference and cognitive structure of Italian political parties", Italian J. of Psych., 2 (1975), 391-402.

(15) TAKANE Y., "Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment", Japanese Psych. Res., 22 (1980), 188-196.

(16) THURSTONE L.L., "A law of comparative judgment", Psych. Rev., 34 (1927), 273-286.

(17) TVERSKY A. and RUSSO J.E., "Substitutability and similarity in binary choice", J. of Math. Psych., 6 (1969), 1-12.

(18) TVERSKY A. and SATTATH S., "Preference trees", Psych. Rev., 86 (1979), 542-573.