Marketing data analysis by dual scaling*

Shizuhiko NISHISATO
Wolfgang GAUL

A survey of various aspects of dual scaling and an assessment of variants of this method with respect to its ability to handle marketing data are presented. The paper looks at dual scaling as a distinct method among those techniques based on the same principle, in particular, as correspondence analysis, for quantifying qualitative data from marketing research.

Data from advertising research are used to illustrate the potentials of dual scaling and some of its important characteristics.

The paper is intended to serve as a useful guide for applications of dual scaling to problems in marketing research.

1. Introduction

1.1. Historical background

In 1860, Gustav Fechner published his classical work, entitled 'Elemente der Psychophysik'. Since then, quantification of human judgement has been an important topic in the social sciences, particularly in psychology. It was not until 1927, however, that a statistical foundation for 'scaling' was established by Leon L. Thurstone in his famous article on the law of comparative judgement (Thurstone (1927)). This paper marked the beginning of the unidimensional approach to quantification of human judgement.

In 1901, Karl Pearson published an epoch-making article on linear combinations of variables as defined in the principal plane (Pearson (1901)). His basic ideas and formulation were then developed into 'principal component analysis' by Harold Hotelling in his 1933 paper. This study made it possible to represent data in multidimensional space and, hence, may be regarded as a marker of the beginning of the multidimensional approach to quantification of human judgement.

With these two distinct streams of major precursors, numerous methods for analyzing human judgement have been proposed during the past several decades. Especially, during the past 55 years or so, a number of so-called 'quantification methods' have been developed (see e.g. the bibliography of Nishisato (1986) and the assessement of different methods for quantifying categorical multivariate data by Tenenhaus and Young (1985)). Although these techniques as referred to by such names as correspondence analysis, homogeneity analysis and dual scaling have a common ground or starting point – i.e., the singular-value or Eckart–Young decomposition – research has passed the stage of 'basic formulation', moving into the phase of own advancements. This phenomenon of branching out is likely to continue, and it is possible that we may see the day when dual scaling, for example, can no longer be used as synonymous as correspondence analysis (see e.g. Nishisato (1978) who showed that his formulation of dual scaling of paired comparisons is an alternative to formulations proposed by Guttman (1946), Slater (1960), Tucker (1960) and Carroll (1972) some of which are referred to as vector models, and Böckenholt and Gaul (1986,

* Research has been supported by the Deutsche Forschungsgemeinschaft, and was performed while the first author was on leave from the Ontario Institute for Studies in Education (OISE), and the University of Toronto, Canada.
Authors' address: Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe, FRG.
The authors would like to thank two anonymous referees for their comments which improved an earlier version of this paper.

Intern. J. of Research in Marketing 5 (1988) 151–170
North-Holland

1987) for probabilistic generalizations and comparisons of vector and ideal point models).

At the present stage of development, dual scaling is ready to be used for the analysis of various kinds of qualitative data. (Dual scaling, as applied to contingency tables, sorting data and multiple-choice data, is nothing but principal component analysis of categorial data; dual scaling, as applied to paired comparisons, ranking and successive categories data, employs some ideas of Thurstonian scaling methods.) A main orientation of research on dual scaling has been and still is to extend applicability of the method to a wider variety of qualitative data, using the so-called dual relations or duality. As is well known, correspondence analysis is not applicable to a data table which contains negative numbers. One of the distinct aspects of dual scaling, however, lies in its extension to a data table which contains negative numbers. This will be discussed later in more detail.

One of the most useful variants of dual scaling is that of forced classification. This is an analogue of discriminant analysis for multiple-choice, paired comparisons, ranking and sorting data, a very simple, yet effective procedure, which can easily be incorporated into a computer program. Its generalized version offers a variety of procedures for market segmentation, conditional analysis and robust quantification.

Within a marketing context the present study reviews and explains in a clear and integral way some developments in dual scaling published mainly by Nishisato and his associates. Main reasons for this exposition are first, the fact that dual scaling is not well known in marketing research, and second, the conviction that dual scaling has far-reaching potential for elucidating complex judgemental data. Up to now, only a few expository papers on essentially the same background methodology as for dual scaling have appeared in the marketing research literature. The phrase 'essentially the same' has important implications for the following discussion, and will be clarified shortly.

1.2. Marketing studies related to dual scaling

From the previous section, it may be clear that there are some methods which are related to, but not the same as, dual scaling. In a bibliography on quantification of categorial data covering the time span 1975–1986 (Nishisato (1986)), one can find some publications of this general quantification method as applied to marketing research. To name only a few examples, see Carroll, Green and Schaffer (1986), Franke (1983, 1985), Gopalan (1986), Hoffman and Franke (1986), Krier and Jackson (1984), Vasserot (1976), and Weingarden and Nishisato (1986). Out of these articles, widely cited papers by Franke (1985) and Hoffman and Franke (1986) are expository of existing methods. For a thorough discussion of the general methodology of correspondence analysis and dual scaling, therefore, one should refer to Benzecri et al. (1973), Greenacre (1984), Lebart, Morineau and Warwick (1984), and Nishisato (1980).

It seems that dual scaling and its competitors are no longer what Hill (1974) called the neglected multivariate methods but are recognized to be of special interest to investigators in marketing research. Considering the complexity of marketing data and the fact that techniques which are geared to multidimensional analyses of qualitative data are needed, dual scaling can serve as one of the versatile alternatives in data analysis.

In the next section the basic formulation of the method will be given. Then, data from advertising research will be used to illustrate the potentials of dual scaling and some of its important characteristics.

Additionally, sample calculations are provided in an appendix to explain how dual scaling is able to handle different types of data and may serve as a useful guide for applications of dual scaling to problems in marketing research.

2. Basic formulation

The term 'dual scaling' was proposed by Nishisato because of its generality and lack of ambiguity for a method originally designed to quantify qualitative data by assigning optimal weights to the rows and columns of a data matrix. To describe the basic formulation let us introduce the following notation:

$F = (f_{ij}) \cong$ an $n \times m$ data matrix, f_{ij} being the frequency in cell (i, j) for a contingency table, and 1 or 0 for multiple-choice data,
$g = (g_i) \cong$ the vector of row marginals of F,
$f = (f_j) \cong$ the vector of column marginals of F,
$D_r = \text{diag}(g_i)$, $D_c = \text{diag}(f_j)$ $\cong$ the diagonal matrices with respect to row and column marginals,
$y = (y_i) \cong$ a vector of weights for the rows of F,
$x = (x_j) \cong$ a vector of weights for the columns of F.

Additionally, let us fix the origin and the unit of the quantified data by

$$g'y = f'x = 0, \tag{1}$$

$$y'D_r y = x'D_c x = f_t, \tag{2}$$

where f_t is the sum (total) of the elements of F, i.e., $f_t = 1'g = 1'f$ (see section A1 in the appendix for computational examples).

The task of quantification is to determine y and x in a certain optimal way. Some of the popular criteria, used during the past 55 years of the history of this methodology, are

(a) Determine y so as to maximize the between-column discrimination.
(b) Determine x so as to maximize the between-row discrimination.
(c) Determine y and x so that the correlation between responses weighted by y and those weighted by x, ρ, is a maximum, where

$$\rho = y'Fx / \left[y'D_r y x'D_c x \right]^{1/2}. \tag{3}$$

(d) Determine y and x so that regression of x on y and regression of y on x be simultaneously linear.
(e) Determine x so as to maximize the internal consistency reliability of y.

This list can be continued. However, the important point is that all these optimization criteria are met by the Eckart–Young or singular-value decomposition of data matrix F, which can be written in the form of dual relations as

$$\rho y = D_r^{-1} Fx \quad \text{and} \quad \rho x = D_c^{-1} F'y, \tag{4}$$

where ρ is a singular value and identical to ρ in eq. (3).

This set of formulas is used practically by all the variants of the methodology mentioned in this paper.

In dual scaling, its extensions to other types of qualitative data than those mentioned in the description of F are achieved by application of eq. (4) to modified data matrices. In other words, data matrix F is replaced by another matrix, appropriate for a specific type of data, D_r and D_c being also redefined accordingly. This makes it possible for dual scaling to handle many types of data. A few of such modifications are now illustrated, using real data.

3. Example: Analysis of print ads for French cognac

In this section, dual scaling is applied to a data set from advertising research to illustrate some of its unique aspects. The data are from Gaul and Böckenholt (1987), who have analyzed paired comparisons data involving cognac print ads and rating data on seven attributes of the print ads, obtained from 69 subjects. The subjects were participants in courses of continued education at the Chamber of Industry and Commerce of Karlsruhe and students enrolled in an introductory course on marketing at the University of Karlsruhe.

In the first part of data collection, the subjects were shown pairs of print ads of French cognac and were asked 'Which of the two ads is more appealing to you?'. There were five brands – Bisquit, Courvoisier, Hennessy, Martell and Remy – and two variants of print ads for each brand designated as brand (1) and brand (2), respectively; hence, 45 pairs of print ads in total.

In the second part of data collection, the subjects were asked to rate on a five-point scale each of the ten print ads with respect to seven attributes, namely, 'credible', 'extravagant', 'imaginative', 'meaningful', 'precious', 'stimulating' and 'symphathetic', which had been pretested in Böckenholt and Gaul (1984) and Gaul (1984).

The question of how advertising messages influence a target audience has brought about an immense amount of literature which will not be surveyed here (see e.g., Gaul and Böckenholt (1987) for an overview). For so-called imagery products – examples are cigarettes, cognacs, and cosmetics – which do not differ in a really significant way with respect to objective brand characteristics, subjective and more consumer perceptions influencing features, such as prestige of ownership, styling and other emotional meanings, are of special importance for advertising. The design of such ads often shows brand name and/or picture of the brand together with a headline (slogan, short text) where the major part is reserved for a picture which induces emotional feelings. More and more advertising uses such kinds of emotional brand differentiation, thus, techniques to evaluate such types of advertising messages are of importance.

In the following sections, it will be shown how dual scaling as a unified approach can be used to analyze the above mentioned perception and preference data from advertising research.

3.1. Paired comparisons data

The 69×45 matrix F of paired comparisons (see e.g., Gaul and Schader (1987) for an explicit presentation of F and a different analysis via aggregations of relations) is used as a starting point for the analysis via dual scaling.

When presenting dual scaling of paired comparisons data we have to consider two relevant matters: analysis of paired comparisons data in general and quantification of data with negative elements. As can be seen in extensive bibliographies by Davidson and Farquhar (1976) and Nishisato (1977, 1978), paired comparisons data are one of the most thoroughly investigated data types. The traditional method of paired comparisons, whether it employs the normal-response model, the angular-response model, or the logistic (Berkson–Terry–Luce) model, relates the difference between two scale values to the probability that one stimulus is preferred to the other (see e.g., Bock and Jones (1968)). A statistical test for goodness of fit of such a model is available, and one can set a confidence interval on a contrast of scale values. Thus, the traditional approach is statistically rigorous and generally preferred to other alternatives. From the practical point of view, however, one must note that it is a unidimensional model and cannot handle individual differences. The fact that the model fits the data indicates reproducibility of paired comparisons proportions from the derived scale values, and not of the responses of each subject. Individual differences in using judgemental criteria (e.g., the attributes 'credible', 'extravagant', 'imaginative', 'meaningful', 'precious', 'stimulating' and 'sympathetic' mentioned earlier) contribute to a multidimensional configuration of stimulus points, and a unidimensional scale obtained by ignoring individual differences can be misleading from the marketing point of view. Dual scaling offers multidimensional analysis of paired comparisons data and assesses contributions of subjects to each dimension. The scale values and subjects' weights on dimensions together contain information to reproduce each subject's paired comparisons responses, not a group statistic such as proportion as it is the case with the traditional unidimensional approach (see also Böckenholt and Gaul (1984, 1986) for additional arguments to perform multidimensional analyses of paired comparisons data).

Another point worth mentioning explicitly is the ability of dual scaling to accommodate a data table with negative elements into its repertory. This extension has enabled dual scaling to handle such additional types as paired comparisons, rank-order and rating data.

Let us consider a small example which is presented here because it may also serve for clarification with respect to the following sections on rating and rank order data.

Example: Dual scaling methodology expects that an element $(i, (j_1, j_2))$ of a paired comparisons matrix is coded as 1 if subject i prefers the first stimulus j_1 of the pair (j_1, j_2), -1 if the second stimulus j_2 is chosen, and 0 for a tied response. Assume that two subjects are involved in a paired comparisons task on three stimuli. Let F be given by

Subjects	Pairs of stimuli $\binom{1}{2}$	$\binom{1}{3}$	$\binom{2}{3}$
1	1	1	−1
2	1	−1	−1

and let x_1, x_2 and x_3 be unknown weights (i.e., scale values) of the three stimuli. In terms of the unknown scale values, the data indicate that $x_1 > x_2$, $x_1 > x_3$, and $x_3 > x_2$ for subject 1, and $x_1 > x_2$, $x_3 > x_1$, and $x_3 > x_2$ for subject 2, where the symbol '$>$' indicates 'is preferred to'. The task of dual scaling is to determine the scale values in such a way that they are maximally discriminative. For this purpose, it is more convenient to transform the subjects-by-pairs of stimuli table F to a subjects-by-stimuli table E, so that the problem may be stated as that of determining weights for the subjects (rows of E) so as to maximize stimulus discriminability (columns of E).

In constructing such a table E, Nishisato (1978) used a Thurstonian analogue of nonmetric individual differences scaling by post-multiplying matrix F with the well-known design matrix for paired comparisons A which in the underlying example has the form

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix},$$

from which the resultant matrix E is derived by

$$E = FA = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -2 & 0 \\ 0 & -2 & 2 \end{bmatrix}.$$

E is also called dominance matrix because a matrix coefficient e_{ij} of E describes a value which can also be obtained by counting the number of times subject i prefers stimulus j to the remaining stimuli minus the number of times where stimulus j is not preferred by subject i.

This matrix E is subjected to dual scaling with 'modified' diagonal matrices D_c and D_r, due to negative elements. To explain this, note that the initial coding of 1, -1 or 0 means the direction of pairwise contrasts. Thus, the paired comparisons data can be expressed in terms of such contrasts with correct directions as follows:

Subjects	Contrasts of scale values		
1	$(x_1 - x_2)$	$(x_1 - x_3)$	$-(x_2 - x_3)$
2	$(x_1 - x_2)$	$-(x_1 - x_3)$	$-(x_2 - x_3)$

and rearranging of the contents of the above table into a subjects-by-stimuli table yields

Subjects	Stimuli 1	2	3
1	x_1, x_1	$-x_2, -x_2$	$-x_3, x_3$
2	$x_1, -x_1$	$-x_2, -x_2$	x_3, x_3

Notice that each cell of this table consists of a number of quantities, which is equal to the number of stimuli minus one. This is how the diagonal matrices, D_c and D_r, are defined when the corresponding data table contains negative elements (D_c is diag((number of subjects) × (number of stimuli minus one)), D_r is diag((number of stimuli) × (number of stimuli minus one)).) When the above table is presented, some elements are cancelled because of opposite signs, resulting in the following table:

Subjects	Stimuli 1	2	3
1	$2x_1$	$-2x_2$	0
2	0	$-2x_2$	$2x_3$

The coefficients of this matrix are the elements of matrix E, that is of FA, defined above where in the underlying example the connection between the contrasts of the stimuli values and matrix A is given by

$$\begin{bmatrix} x_1 - x_2 \\ x_1 - x_3 \\ x_2 - x_3 \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

In general, when n stimuli and N subjects are used in paired comparisons, each cell of matrix E reflects the outcome of $(n-1)$ comparisons so that the total number of responses of each row of the subjects-by-stimuli dominance matrix is $n(n-1)$ and the total number of responses of each column is $N(n-1)$. Some readers would notice that this formulation is very much like the formulation of the data matrix after removal of the so-called trivial solution. Indeed, under this formulation, too, each eigenvalue is the variance of quantified responses.

To sum up, from the small example we have learned the following:

If N indicates the number of subjects and n the number of stimuli (with $N = 69$ and $n = 10$ in the underlying print ads example) then data matrix F is $N \times n(n-1)/2$. This matrix is converted to an $N \times n$ dominance matrix, $E = (e_{ij})$, where e_{ij} is the number of times subject i prefers stimulus j to the remaining $(n-1)$ stimuli minus the number of times subject i does not prefer stimulus j over the other stimuli.

Simultaneously with the replacement of F by E, the diagonal matrices, D_r and D_c, must be re-defined. For each subject and each object there are $(n-1)$ comparisons, thus, in counting all row and column comparisons, one obtains $D_r = \text{diag}(n(n-1))$, $D_c = \text{diag}(N(n-1))$ and $f_t = Nn(n-1)$ (see section A2 in the appendix for further computational examples).

With the redefined D_r and D_c matrices and the substitution of F by E formula (4) determines the dual scaling solution for the paired comparisons data, namely, weights y and x that maximize ρ.

This extension of the method to a matrix with negative elements may appear extraordinary, casting a doubt about the dual relations. Notice, however, that the dual relations of eq. (4) apply to multiple dimensions, of which all the decompositions, except that of the trivial solution, are associated with matrices containing negative elements. To show this point, it is known that eq. (4) can be rewritten in a multidimensional form as

$$F = D_r(11' + \rho_1 y_1 x_1' + \rho_2 y_2 x_2' + \dots + \rho_k y_k x_k')D_c/f_t.$$

Dual relations exist with respect to all the following matrices:

$F,\ F - D_r 11' D_c/f_t,$

$F - D_r(11' + \rho_1 y_1 x_1')D_c/f_t,$

$F - D_r(11' + \rho_1 y_1 x_1' + \rho_2 y_2 x_2')D_c/f_t,$

and so forth.

Yet, D_r and D_c are defined in terms of the row and the column marginals of matrix F, respectively, because the above residual matrices are all based on so many numbers of responses as given by the marginals of matrix F. Dual scaling of paired comparisons data amounts to an extension of the method to a situation in which the so-called trivial solution is already removed from the input matrix for scaling.

The first five dimensions of the evaluation of the underlying data account for about 86% of the total variance, with individual contributions being respectively 36% (dimension I), 19% (dimension II), 14% (dimension III), 9% (dimension IV) and 7% (dimension V). Fig. 1 shows the configuration of the print ads in the first two major dimensions, which account for about 56% of the total variance.

The results are very similar to those obtained by Gaul and Böckenholt (1987) using a probabilistic ideal point model. This similarity in results was more or less anticipated because the two approaches can be considered as based on the generalized (weighted) least-squares principle. The two approaches should not differ very much when the first

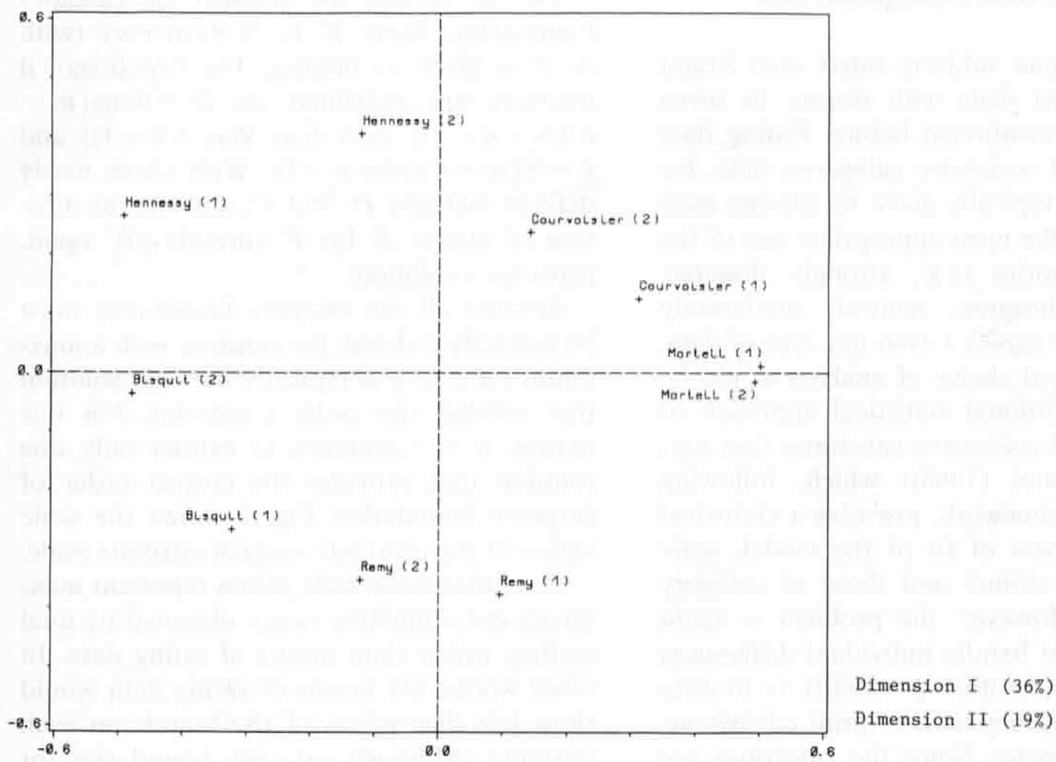

Fig. 1. Analysis of paired comparisons data for ten print ads.

one or two dimensions are looked at, but the discrepancy between them would widen as one would go down to less dominant dimensions. As for the first dimension, it is known that even the means of the dominance scores resemble those optimal results from dual scaling. Considering that dimension I is dominant, the real winners are Martell (1) and Martell (2). It is interesting to note in fig. 1 that the two print ads of each brand are closely located or clustered. Does this also mean that the two ads of each brand have similar attributes? In addition, one may ask what attributes appear to contribute to the popularity of the Martell print ads, as opposed to Hennessy (1) and Bisquit (2), in dimension I, and also what attributes distinguish e.g. between the Hennessy print ads and the Remy print ads in dimension II. In the next section, we will try to find answers to these questions.

3.2. Rating (successive categories) data

The sixty-nine subjects rated each brand on a five-point scale with respect to seven attributes, as mentioned before. Rating data are also called successive categories data, for responses are typically given by placing each stimulus into the most appropriate one of the ordered categories (e.g., strongly disagree, moderately disagree, neutral, moderately agree, strongly agree). Given this type of data, the most natural choice of analysis would be again the traditional statistical approach of the method of successive categories (see e.g., Bock and Jones (1968)) which, following Thurstone's framework, provides a statistical test for goodness of fit of the model, scale values of the stimuli and those of category boundaries. However, the problem is again the inability to handle individual differences in judgement. Another possibility is to construct a contingency table of print ads-by-categories frequencies. Since the categories are ordered, it would be preferred to impose a complete-order constraint on the weights for the categories. Quantification under constraints concerning partial or complete orders can be carried out with any of the existing methods (see e.g. Nishisato and Arri (1975), de Leeuw, Young and Takane (1976), Tanaka and Kodake (1980), and Nishisato (1980)). Such an approach, however, does not provide estimates of category boundaries, which are sometimes useful, nor taken into consideration any individual differences in judgement. Thus, it looks as though dual scaling may be preferred to these alternatives. Let us now turn to the dual scaling formulation.

By taking into consideration category boundaries between the different categories, rating data can be converted to rank order data of stimuli and category boundaries, which are then transformed into the elements of a subject-by-(category boundaries, stimuli) matrix E of dominance scores (see section A3 in the appendix for computational examples).

Let m denote the number of category boundaries. Then E is $N \times (m+n)$ (with N, n as given as before). The two diagonal matrices are redefined as $D_r = \mathrm{diag}((m+n)(m+n-1))$, $D_c = \mathrm{diag}(N(m+n-1))$, and $f_t = N(m+n)(m+n-1)$. With these newly defined matrices D_r and D_c and the substitution of matrix E for F, formula (4), again, provides a solution.

Because all the category boundaries must be correctly ordered, the solution with a maximum value of ρ is typically the only solution that satisfies this order constraint. For this reason, it is customary to extract only one solution that provides the correct order of category boundaries. Fig. 2 shows the scale values of the print ads on each attribute scale.

Note that those scale values represent maximally discriminative values obtained by dual scaling, rather than means of rating data. In other words, the means of rating data would show less dispersions of the brands on each attribute. Although category boundaries for each attribute scale were also calculated as

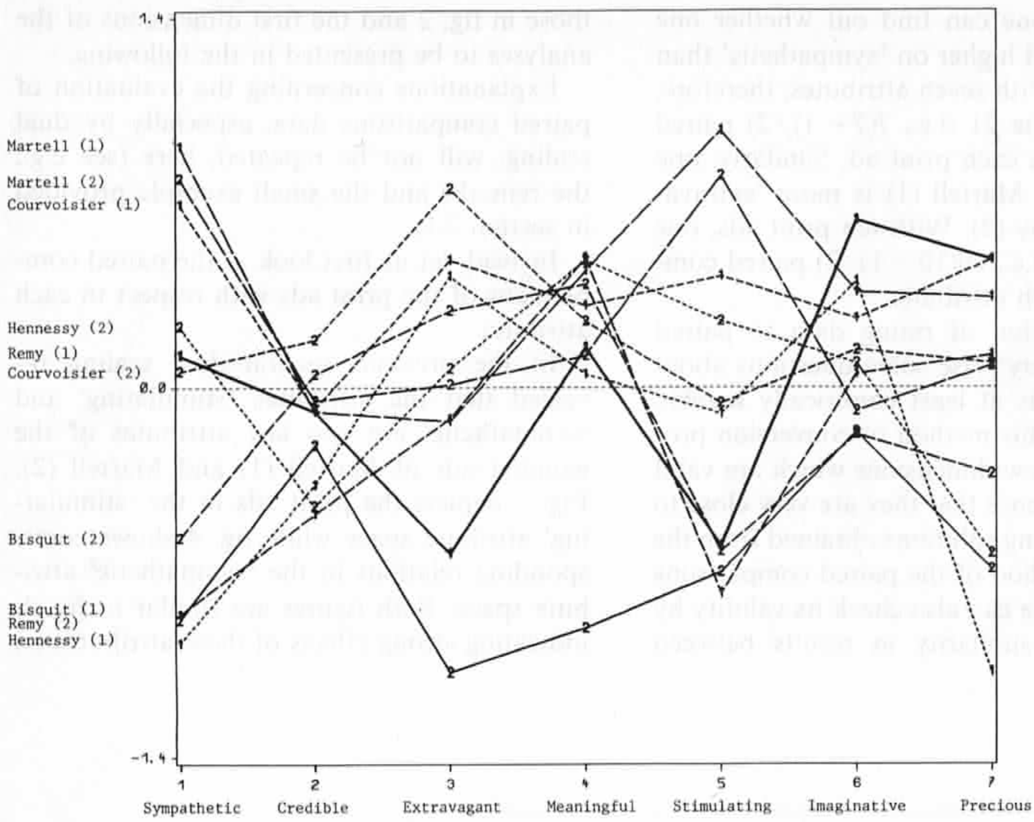

Fig. 2. Scale values for the print ads on seven attribute scales.

part of the dual scaling output, they are not shown in fig. 2 to simplify the graphical representation.

A noteworthy point is that the two attributes 'sympathetic' and 'stimulating' provide wider distributions of the print ads than the others, suggesting a stronger discriminative property with respect to the subjects' judgements of the print ads than others. In contrast, 'credible' and 'imaginative' show much narrower distributions of the print ads.

Note that the winners Martell (1) and Martell (2) are very high on 'sympathetic' and 'stimulating', and that the profiles of the two ads are very similar. The profile of Hennessy (1) is quite different from that of Hennessy (2). Hennessy (1) is least 'stimulating' and least 'precious'. Notice, also, that Courvoisier (1) and Courvoisier (2) are most 'extravagant', and that Remy (1) and Remy (2) are the least 'extravagant'.

The variance among the ten print ads explained by each unidimensional attribute scale in fig. 2 is – from the largest to the smallest – as follows: 'stimulating' (47%), 'sympathetic' (42%), 'credible' (41%), 'precious' (41%), 'extravagant' (39%), 'meaningful' (34%), and 'imaginative' (34%). These numbers suggest multidimensionality of the data.

For multidimensional analysis, it is a common procedure of dual scaling to convert rating data to paired comparisons data.

3.3. *Paired comparisons data generated from rating data*

When rating data are available, one can generate paired comparisons data from them.

For instance, one can find out whether one print ad is rated higher on 'sympathetic' than on 'credible'. With seven attributes, therefore, one can generate 21 (i.e., $7(7-1)/2$) paired comparisons on each print ad. Similarly, one can ask if e.g., Martell (1) is more 'extravagant' than Remy (2). With ten print ads, one can derive 45 (i.e., $10(10-1)/2$) paired comparisons on each attribute.

This conversion of rating data to paired comparisons may raise some questions about its validity. It is at least empirically known, however, that this method of conversion provides the first few dimensions which are valid – valid in the sense that they are very close to the corresponding solutions obtained from the traditional method of the paired comparisons experiment. One can also check its validity by observing the similarity in results between those in fig. 2 and the first dimensions of the analyses to be presented in the following.

Explanations concerning the evaluation of paired comparisons data, especially by dual scaling, will not be repeated, here (see e.g., the remarks and the small example provided in section 3.1).

Instead, let us first look at the paired comparisons of the print ads with respect to each attribute.

In the previous section dual scaling revealed that the attributes 'stimulating' and 'sympathetic' are two key attributes of the winning ads of Martell (1) and Martell (2). Fig. 3 depicts the print ads in the 'stimulating' attribute space while fig. 4 shows corresponding relations in the 'sympathetic' attribute space. Both figures are similar to fig. 1, indicating strong effects of these attributes on

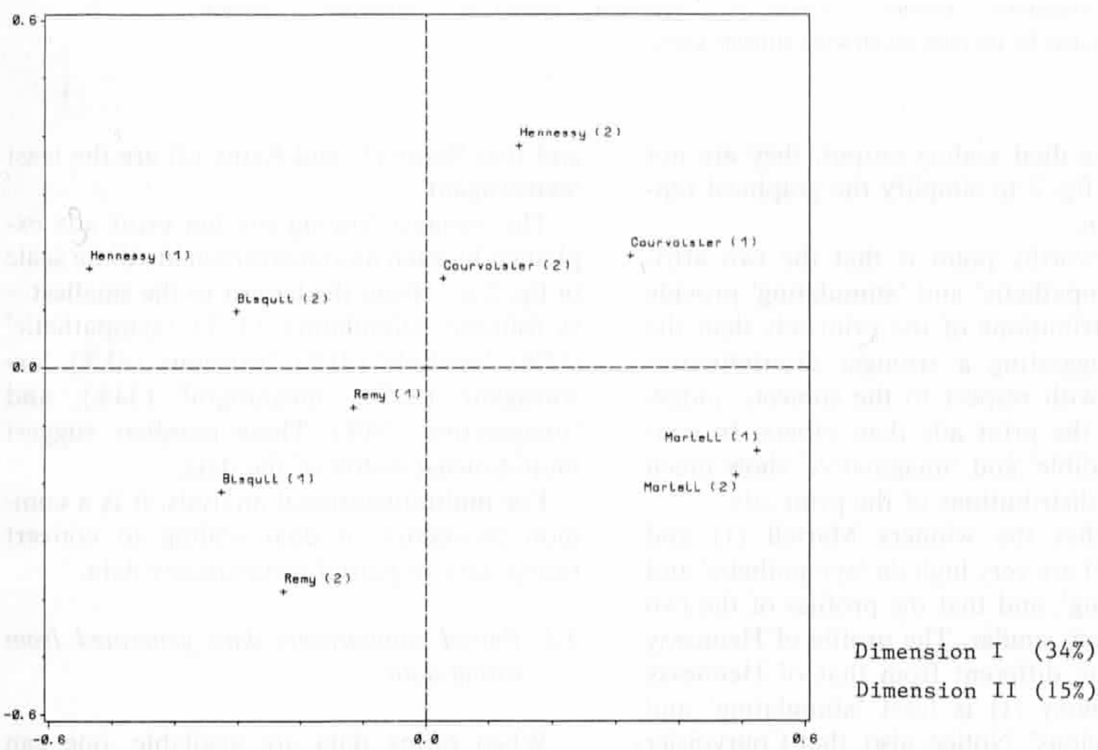

Fig. 3. Representation of the print ads based on the attribute 'stimulating'.

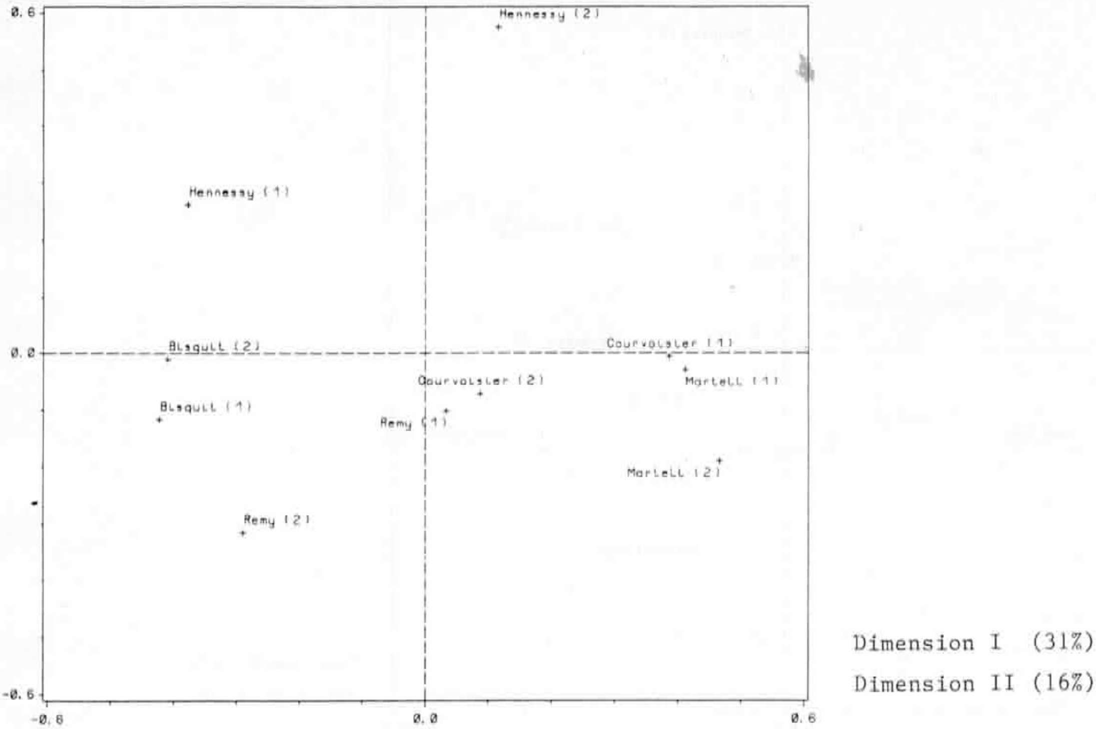

Fig. 4. Representation of the print ads based on the attribute 'sympathetic'.

determining the over-all paired comparisons structure. The respective dimension I of the current results of the two attributes (the projection of the locations of the print ads on this dimension) resembles the relations in fig. 2 very closely, suggesting the validity of the conversion procedure of rating data into paired comparisons data.

Since there are so many results to look at, let us choose only more attribute. Fig. 5 shows the configuration of the print ads in the 'extravagant' attribute space. Dimension I of fig. 5, again, supports the corresponding results of fig. 2. Note, however, that now the two ads of each brand show greater differences in this attribute space than in the spaces of the more decisive attributes of 'stimulating' and 'sympathetic'.

On the other side we can (re)examine what attributes, if any, seem to be key characteristics of each ad of each brand, using the paired comparisons data on the seven attributes, generated from the rating data. When the ten print ads were analyzed with respect to the attributes, it was discovered, without any exceptions, that the two attributes, 'meaningful' and 'imaginative', formed a tight cluster at one end of a principal axis with the rest of the attributes distributing in the other side of the space. Since such 'common' findings does not contribute to distinct characterizations of the print ads, it was decided to discard those two attributes, and to investigate characterizations of each print ad in terms of the other five attributes. Fig. 6 shows the attribute structures of Martell (1) and Martell (2). The distributions of the attributes are, as expected, very similar between the two print ads Martell (1) and Martell (2), and the two are preferred over the other ads because of the attributes 'stimulating' and 'sympathetic', as opposed to 'extravagant', 'precious' and 'credible' (also some subjects preferred the latter three attributes to the first two as one

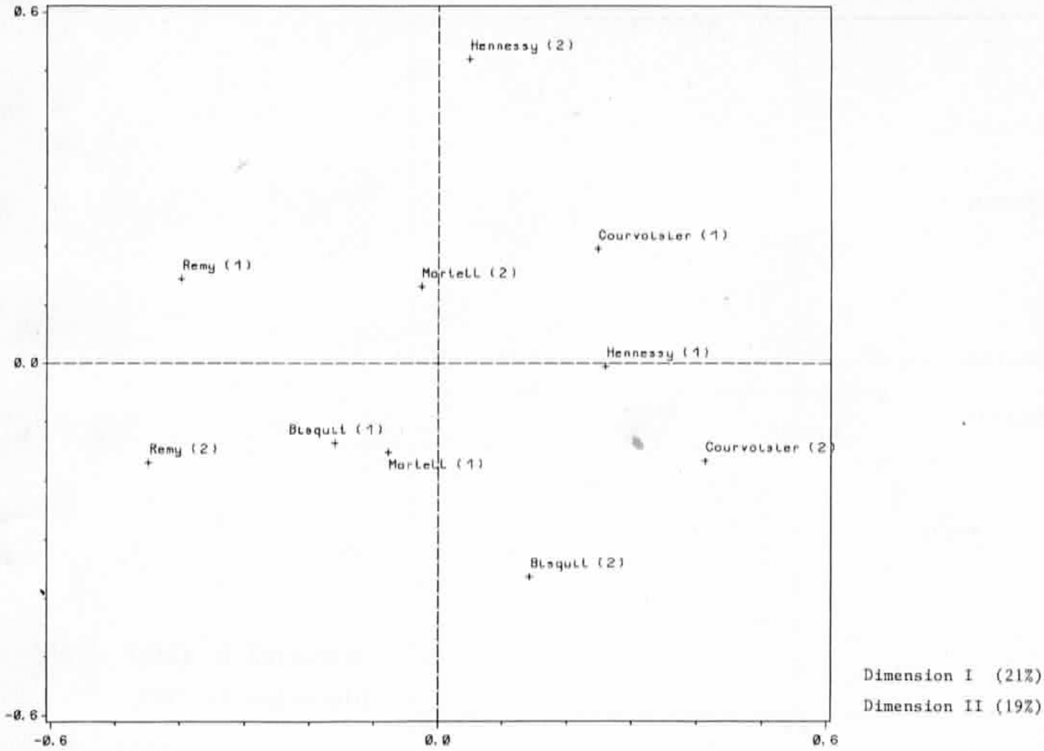

Fig. 5. Representation of the print ads based on the attribute 'extravagant'.

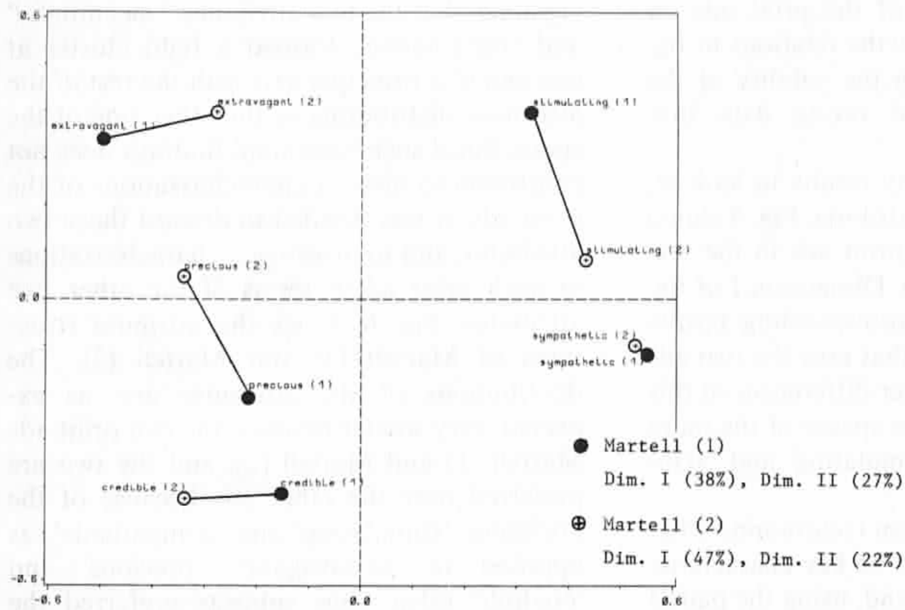

Fig. 6. Distribution of attributes of the Martell ads.

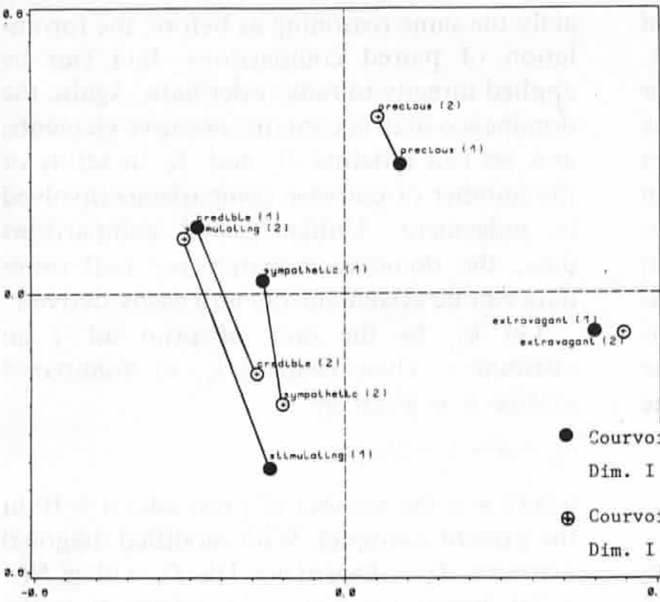

Fig. 7. Distribution of attributes of the Courvoisier ads.

can infer from negative weights of those subjects on dimension I which may give rise to segmentation of subjects). Dimension II also splits subjects into two groups, those who like Martell's attributes 'extravagant' and 'stimulating', and those who like the attribute 'credible'.

Fig. 7 shows the attribute structures for the Courvoisier ads, and was chosen as an example which depicts stronger differences with respect to some attributes. Courvoisier ads are preferred over other ads because of the attribute 'extravagant' which supports findings from e.g. fig. 2. The attribute 'precious' determines part of the meaning of dimension II while attributes as 'credible', 'sympathetic' and 'stimulating', which are clearly opposite to 'extravagant' on dimension I, are very different for the two ads with respect to dimension II.

Such findings, available for all print ads, support efforts to evaluate key characteristics for the ads of the different brands.

The above data can be rearranged into a frequency table with the number of times each category of each attribute is associated

Table 1
Ranking of the print ads on successive category scales of attributes.

Attributes	Print ads									
	1 [a]	2	3	4	5	6	7	8	9	10
Sympathetic	5	10	3	8	1	9	4	6	7	2
Credible	5	8	4	10	6	7	9	1	2	3
Extravagant	9	3	2	7	8	10	5	1	4	6
Meaningful	7	9	5	1	3	10	2	6	4	8
Stimulating	6	10	3	9	1	8	4	5	7	2
Imaginative	1	2	4	7	3	9	6	5	10	8
Precious	1	10	2	4	3	8	8	5	7	6

[a] *Notes:* 1 = Remy (1), 2 = Hennessy (1), 3 = Courvoisier (1), 4 = Bisquit (1), 5 = Martell (1), 6 = Remy (2), 7 = Hennessy (2), 8 = Courvoisier (2), 9 = Bisquit (2), 10 = Martell (2).

with each print ad. The frequency table of categories of attributes by (print ads × attributes) may be scaled with the weak order constraint on the categories. This analysis may be worth carrying out. However, this was not done here for the reasons that first of all even under the weak order constraint analysis provides only one solution within the current formulation of dual scaling, and that the positions of the attributes and the print ads cannot be isolated since the columns of the juxtaposed matrix are combinations of the print ads and attributes.

3.4. Rank order data

In the analysis of section 3.2 dual scaling of the successive categories data provided unidimensional scales of the ten print ads on each of the seven atttributes. To show an application of dual scaling to rank order data, these scale values of the ads were ranked within each attribute, leading to a 7 × 10 matrix of rank order data; see table 1. There are many methods to analyze rank order data (see e.g., Guilford (1954) for an early reference and the bibliographies by Nishisato (1977, 1978) for more recent developments). As is known, the subjects-by-stimuli table of rank orders is ipsative and row-conditional. In other words, direct comparisons across subjects are meaningless and principal component analysis of stimuli, for example, would produce only artefacts of ipsativity. The stimuli-by-ranks table of frequencies has even tighter constraints than the subjects-by-stimuli table, namely, the constant row marginals and the constant column marginals. Bock and Jones (1968), for example, discuss a statistical treatment of the table, following Thurstone's treatment of rank orders (Thurstone (1931)) using a unidimensional model. We can advance here the same argument as that mentioned in relation to paired comparisons data. We wish to capture individual differences and multidimensionality of stimuli. Following exactly the same reasoning as before, the formulation of paired comparisons data can be applied directly to rank order data. Again, the dominance matrix contains negative elements, and we can redefine D_c and D_r in terms of the number of pairwise comparisons involved in judgement. Unlike paired comparisons data, the dominance matrix for rank-order data can be systematically and easily derived.

Let k_{ij} be the rank of print ad j on attribute i. Then, element e_{ij} of dominance matrix E is given by

$$e_{ij} = n + 1 - 2k_{ij},$$

where n is the number of print ads ($n = 10$ in the present example). With modified diagonal matrices, $D_r = \text{diag}(n(n-1))$, $D_c = \text{diag}(N(n-1))$, hence $f_t = Nn(n-1)$, where N is the number of attributes ($N = 7$ in the present example), dual scaling can be carried out.

The rationale for defining D_r and D_c is the same as given for paired comparisons data. Because of limited space, only a brief summary of main results will be given; see table 2.

Three solutions were obtained. It is interesting to note that dual scaling of a matrix consisting of seven independently obtained dimensions provides results similar to some of the earlier multidimensional results carried out in previous single analyses.

3.5. Forced classification application

Rating data with discrete categories can be treated as multiple-choice data, where the number of categories corresponds to the number of response options. In the present example, the rating data on the attributes of the print ads can be analyzed as multiple-choice data.

Suppose that N subjects answered n multiple-choice questions, and that the data are expressed as

$$F = [F_1, F_2, \ldots, F_j, \ldots, F_n],$$

Table 2
Main contributors to three solutions from rank orders of scale values.

Solution	Contributing attributes	Major patterns
1	Sympathetic, Stimulating,	[Martell (1), Courvoisier (1), Martell (2)] *versus*
(39%)	Precious	[Hennessy (1), Remy (2)]
2	Credible	[Courvoisier (2), Bisquit (2), Martell (2)] *versus*
(21%)	Extravagant	[Bisquit (1), Martell (1), Remy (1)]
3	Imaginative	[Remy (1), Hennessy (1)] *versus*
(16%)	Meaningful	[Bisquit (2), Hennessy (2)]

where F_j is $N \times m_j$, consisting of 1's (responses) and 0's (no responses), m_j being the number of options of item (question) j. Matrix F_j is therefore the incidence matrix or response-pattern matrix of item j.

Suppose that F_j is multiplied by a positive constant k, and that the resultant matrix is denoted as $F(j \times k)$, i.e.,

$$F(j \times k) = [F_1, F_2, \ldots, kF_j, \ldots, F_n].$$

As k approaches infinity, it is known (Nishisato (1984)) that dual scaling of $F(j \times k)$ becomes identical to that of the original matrix projected onto the subspace spanned by the columns of F_j, that is, dual scaling of P_jF, where $P_j = F_j(F_j'F_j)^{-1}F_j'$ (see section A4 in the appendix for computational examples).

If question j is about age, dual scaling of P_jF quantifies the data so as to discriminate optimally between young and old, which therefore enables the investigator to identify questions that contribute to distinguishing between subgroups or segments of subjects. In the above example, age is referred to as the forced classification criterion.

This idea can be extended to dual scaling of paired comparisons data in which a pair is chosen as the criterion. This mode of forced classification generates a scale that maximizes the difference between the two stimuli in the criterion pair, that is, a bipolar scale with the two stimuli occupying the two ends of the continuum. This form of forced classification is used in this study. Gaul and Schader (1987) applied a 'CAR (Clusterwise Aggregation of Relations)' technique to the same paired comparisons data on the ten print ads. Although their method is at least theoretically very different from dual scaling, it is, of course, of interest to compare the two results. Gaul and Schader (1987) provided two optimal segments among the 69 subjects and presented the corresponding two distinct orderings of the print ads, together with a graph based on a segmentation approach of the Gaul and Böckenholt (1987) probabilistic ideal point model.

The first segment by the Gaul–Schader algorithm ranked Martell (1) and Bisquit (2) at the two extreme positions. Therefore, forced classification of dual scaling was carried out with Martell (1) and Bisquit (2) as the criterion pair at the two ends of the scale. This ordering is (from the most preferred to the least):

Martell (1) > Martell (2) > Courvoisier (1)
> Courvoisier (2) > Remy (1)
> Hennessy (2)
> Remy (2) > Bisquit (1) > Hennessy (1)
> Bisquit (2).

This ranking is identical with that of the Gaul–Schader method. The difference between the two approaches becomes apparent when the segmentation problem is considered. Unlike the CAR approach, dual scaling assigns weights to the subjects according to their contributions to each dimension, but without producing distinct segments. The 24 subjects of the first segment yielded from the CAR approach have generally large weights in the dual scaling solution, suggesting a high degree of comparability between the two results.

One of the differences between the CAR and dual scaling approaches can be regarded as that between disjoint and discrete clustering and continuous clustering. Most techniques of cluster analysis generate disjoint clusters, where an element is classified into a single cluster. In contrast, dual scaling, principal component analysis and factor analysis assign a weight to each variable for each dimension (cluster), and such a weight provides information as to how many dimensions a particular variable is contributing to. In the latter case, the problem of cutting points in deciding which dimension a variable should be associated with becomes a serious issue to investigate this being the case if one wants to use the technique for clustering variables. It is not surprising that the two approaches have shown similar results on one dimension, but it is natural that they become different in subsequent dimensions.

Dual scaling provides a very clear-cut method for segmentation only when an external criterion exists.

3.6. *Résumé of results on the print ads data*

When the dual relations as given in formula (4) are applied to a data matrix F and the diagonal matrices D_r and D_c (or appropriate redefined matrices), quite a number of possible and interesting ways of optimal quantifications of qualitative data can be performed. The present paper has illustrated only a few of them.

Using the Gaul and Böckenholt (1987) data, five applications of dual scaling were presented.

When a new advertising campaign is needed or when the effect of own print ads in relation to those of main competitors in an underlying product class has to be assessed, an evaluation of data as described in the previous sections can be useful.

Different analyses found the Martell print ads to be the winners with respect to the first dimension as in fig. 1. The attributes 'stimulating' and 'sympathetic' are most important for their perceptual positioning as can be seen from fig. 2. The analyses as depicted in figs. 3 and 4 relate the print ads with respect to these important attributes (stimulating, sympathetic) in a paired comparisons based multidimensional representation while fig. 5 shows such a representation based on the attribute 'extravagant'. On the other side, for each print ad (or for the two print ads of the same brand) one can represent the attributes (or the differences of the attribute positions with respect to the two print ads of the same brand) in multidimensional spaces, see figs. 6 and 7, to assess their contributions to the main perceptual dimensions. Also, rank orders can help to reveal patterns which support the characterization of the print ads in terms of those attributes which cause major patterns.

Finally, the forced classification variant of dual scaling was applied to paired comparisons data with a pair of objects as the forced classification criterion. The present study has demonstrated only several potential applications of dual scaling to marketing data. For further findings concerning advertising research aspects, see Gaul and Böckenholt (1987) and Gaul and Schader (1987).

4. Further remarks and conclusions

Since the main purpose of the present paper is in a brief exposition of some of the unique aspects of dual scaling, detailed descriptions of the different variants were not provided.

With respect to each aspect, interested readers are referred to the original articles (Nishisato (1978) for paired comparisons and rank order data, Nishisato (1980) and Nishisato and Sheu (1984) for successive categories (rating) data, and Nishisato (1984, 1987) for forced classification). Other topics not discussed here, as treatment of ordered categories, analysis of multiway data matrices and analysis of sorting data are illustrated in

Nishisato (1980) and Nishisato and Nishisato (1984).

One can also consider dual scaling of categorized data of originally continuous data. Thus, it looks as if one can apply dual scaling to a large number of situations. It is hoped that the present paper may serve as a useful introduction to the potential of this methodology in marketing data analysis.

Appendix

Sample calculations are provided here to explain how dual scaling is able to handle different types of marketing data.

A1. Basic formulation

Consider the following 2×3 contingency table,

$$F = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 7 & 4 \end{bmatrix},$$

where the two rows correspond to the categories of the row variable (e.g., union vs. non-union members) and the three columns to the column variable (e.g., agree, neutral, disagree). Then g, f, D_r, D_c, y and x are given by

$$g = \begin{bmatrix} 10 \\ 14 \end{bmatrix}, \quad D_r = \begin{bmatrix} 10 & 0 \\ 0 & 14 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

$$f = \begin{bmatrix} 8 \\ 10 \\ 6 \end{bmatrix}, \quad D_c = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 6 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

where y and x have to fulfil the constraints

$$10y_1 + 14y_2 = 0, \qquad 10y_1^2 + 14y_2^2 = 24,$$

$$8x_1 + 10x_2 + 6x_3 = 0,$$

$$8x_1^2 + 10x_2^2 + 6x_3^2 = 24,$$

according to formulae (1), (2).

A2. Paired comparisons data analysis by dual scaling

Consider that five subjects participated in a paired comparisons experiment on four stimuli (i.e., $N = 5$, $n = 4$) and that data as shown in the following 5×6 data matrix F were obtained:

Subjects	Pairs of objects $\binom{1}{2}$	$\binom{1}{3}$	$\binom{1}{4}$	$\binom{2}{3}$	$\binom{2}{4}$	$\binom{3}{4}$
1	1	−1	1	−1	1	1
2	1	1	1	1	1	−1
3	1	−1	1	−1	−1	−1
4	−1	1	0	1	1	1
5	0	−1	−1	−1	−1	−1

Then, matrix E is 5×4, with elements e_{ij} calculated as follows. Look at subject 1 and stimulus 1. This subject preferred stimulus 1 over 2, 3 over 1, and 1 over 4 ((1, (1, 2)) = 1, (1, (1, 3)) = −1, (1, (1, 4)) = 1), hence $e_{11} = 1$ (i.e., stimulus 1 is preferred to the other three stimuli twice, and not preferred to the others once). Similarly, for subject 1 and stimulus 4 ((1, (1, 4)) = 1, (1, (2, 4)) = 1, (1, (3, 4)) = 1) one gets $e_{14} = -3$. In this way the dominance matrix is calculated as

$$E = \begin{bmatrix} 1 & -1 & 3 & -3 \\ 3 & 1 & -3 & -1 \\ 1 & -3 & 1 & 1 \\ 0 & 3 & -1 & -2 \\ -2 & -2 & 1 & 3 \end{bmatrix}.$$

More formally, as was demonstrated in the small example in section 3.1, we can also obtain matrix E by postmultiplying F with design matrix

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

constructed from the contrast equation with respect to the underlying situation

$$\begin{bmatrix} x_1 - x_2 \\ x_1 - x_3 \\ x_1 - x_4 \\ x_2 - x_3 \\ x_2 - x_4 \\ x_3 - x_4 \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

As is seen in this example matrix E has the property that every row marginal is zero. Thus, D_r and D_c must be redefined. Since every e_{ij} is based on $(n-1)$ comparisons, in counting the number of comparisons in rows or columns one gets $D_r = \text{diag}(n(n-1))$ and $D_c = \text{diag}(N(n-1)$, so that $f_t = Nn(n-1)$.

In the present example, we have

$$D_r = \begin{bmatrix} 12 & 0 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{bmatrix},$$

$$D_c = \begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 15 & 0 & 0 \\ 0 & 0 & 15 & 0 \\ 0 & 0 & 0 & 15 \end{bmatrix}, \quad f_t = 60.$$

A3. Rating (successive categories) data analysis by dual scaling

Consider an example in which four subjects rated three stimuli in terms of four categories (never (1), sometimes (2), often (3), always (4)) as follows:

Subjects	Stimuli 1	2	3
1	3	2	4
2	2	1	3
3	1	1	2
4	2	4	4

Let us indicate by t_1, t_2 and t_3 three category boundaries, t_1 being between never and sometimes, t_2 between sometimes and often, and t_3 between often and always. Then, e.g., the decision of subject 1 can be rearranged as follows:

t_1 : stimulus 2 : t_2 : stimulus 1 : t_3 : stimulus 3.

In this way, the original data are first transformed to rank order data of both category boundaries and stimuli as described by the following matrix:

Subjects	t_1	t_2	t_3	$St.1$	$St.2$	$St.3$
	Category boundaries and stimuli					
1	1	3	5	4	2	6
2	2	4	6	3	1	5
3	3	5	6	1.5	1.5	4
4	1	3	4	2	5.5	5.5

In a second step, the dominance matrix E is calculated. If we indicate by m the number of category boundaries (and by N respectively n the number of subjects respectively stimuli) then E is $N \times (m+n)$ (with $N = 4$, $m = 3$ and $n = 3$ in this example) and given by

$$E = \begin{bmatrix} 5 & 1 & -3 & -1 & 3 & -5 \\ 3 & -1 & -5 & 1 & 5 & -3 \\ 1 & -3 & -5 & 4 & 4 & -1 \\ 5 & 1 & -1 & 3 & -4 & -4 \end{bmatrix}.$$

Again, D_r and D_c have to be defined suitably (see also section A2). Here, for each subject and each category boundary or stimulus we have $(m+n-1)$ comparisons, hence $D_r = \text{diag}((m+n)(m+n-1))$, $D_c = \text{diag}(N(m+n-1))$, and $f_t = N(m+n)(m+n-1)$. In the present example, D_r is the 4×4 diagonal matrix with 30 in the main diagonal, D_c is the 6×6 diagonal matrix with 20 in the main diagonal, and $f_t = 120$.

A4. The forced classification variant of dual scaling

Consider the following example of multiple-choice data, collected from five subjects on three items:

	Items							
	1			2			3	
Subjects	Options			Options			Options	
	1	2	3	1	2	3	1	2
1	1	0	0	1	0	0	0	1
2	1	0	0	0	1	0	1	0
3	0	1	0	0	0	1	0	1
4	0	1	0	0	1	0	0	1
5	0	0	1	0	0	1	1	0

The above data matrix of 1s and 0s will be indicated by $[F_1, F_2, F_3]$, the subscripts referring to items.

If we assign to the eight options of three items unknown option weights $x_1, \ldots, x_8$, the subjects-by-items matrix of quantified responses is given by

	Items		
Subjects	1	2	3
1	x_1	x_4	x_8
2	x_1	x_5	x_7
3	x_2	x_6	x_8
4	x_2	x_5	x_8
5	x_3	x_6	x_7

Let us define the total score of subject i to be the sum of the three weights of the options the subject chose.

Then, consider product-moment correlation between the total scores and each item, r_{jt}, $j = 1, 2, 3$, and inter-item correlations, r_{jh}, $j \neq h$, $j = 1, 2, 3$, $h = 1, 2, 3$. In the following matrices

$$\text{(a) } [F_1, F_2, F_3], \quad \text{(b) } [F_1, F_2, F_3, F_3],$$

$$\text{(c) } [F_1, F_2, F_3, F_3, F_3],$$

item 3 is repeated twice in (b) and three times in (c). The resultant correlation coefficients of interest are

$$\text{(a) } r_{13} = 0.54, \quad r_{23} = 0.30, \quad r_{3t} = 0.73,$$

$$\text{(b) } r_{13} = 0.74, \quad r_{23} = 0.37, \quad r_{3t} = 0.97,$$

$$\text{(c) } r_{13} = 0.76, \quad r_{23} = 0.40, \quad r_{3t} = 0.99,$$

and one can see that as the number of repetitions of a particular item, item 3 in this example, increases, so do the corresponding correlation r_{j3} and correlation r_{3t}. This is understandable because the incidence matrix of the corresponding item, F_3 in this example, becomes more dominant in the data matrix as the number of repetitions increases. This is the proceeding of forced classification. A special item is chosen and called criterion item. As the number of repetitions of the criterion item increases, geometrically, the criterion incidence matrix eventually becomes a basis for the principal plane. Computationally, it is advantageous to use, for example, $[F_1, F_2, 2F_3]$ for (b) and $[F_1, F_2, 3F_3]$ for (c), for these matrices are structurally equivalent to the original ones of (b) and (c).

References

Benzécri et al., 1973. L'Analyse des Donneés: II. L'Analyse des correspondances. Paris: Dunod.

Bock, R.D. and L.V. Jones, 1968. Measurement and prediction of judgement and choice. San Francisco, CA: Holden-Day.

Böckenholt, I. and W. Gaul, 1984. A multidimensional analysis of consumer preference judgements related to print ads. In: EMAC/ESOMAR (eds.), Methodological advances in marketing research in theory and practice. EMAC/ESOMAR Symposium, Copenhagen: ESOMAR, 83–105.

Böckenholt, I. and W. Gaul, 1986. Analysis of choice behaviour via probabilistic ideal point and vector models. Applied Stochastic Models and Data Analysis 2, 209–226.

Böckenholt, I. and W. Gaul, 1987. Probabilistic multidimensional scaling of paired comparisons data. In: Bock, H.H. (ed.), Proceedings of the First Conference of the International Federation of Classification Societies. Aachen: North-Holland, forthcoming.

Carroll, J.D., 1972. Individual differences and multidimensional scaling. In: Shepard, R.N., A.K. Romney and S.B. Nerlove (eds.), Multidimensional scaling: Theory and applications in the behavioral sciences, Vol. I. New York: Seminar Press.

Carroll, J.D., P.E. Green and C. Schaffer, 1986. Interpoint distance comparison in correspondence analysis. Journal of Marketing Research 23, 271–280.

Davidson, R.R. and P.H. Farquhar, 1976. A bibliography on the method of paired comparisons, Biometrics 32, 241–252.

De Leeuw, J., F.W. Young and Y. Takane, 1976. Additive structure in qualitative data: An alternative least squares method with optimal scaling features. Psychometrika 41, 471–503.

Fechner, G.T., 1860. Elemente der Psychophysik. Leipzig: Breitkopf and Hartel.

Franke, G.R., 1983. Dual scaling: A model for interpreting and quantifying categorial data. In: Darden, W.R., K.B. Monroe and W.R. Dillon (eds.), Research methods and causal modeling in marketing. Chicago, IL: American Marketing Association.

Franke, G.R., 1985. Evaluating measures through data quantification: Applying dual scaling to an advertising copytest. Journal of Business Research 13, 61–69.

Gaul, W., 1984. Datenanalyse auf der Basis von Ordinalurteilen. Studien zur Klassifikation 15, 142–152.

Gaul, W. and I. Böckenholt, 1987. Generalized Thurstonian scaling of advertising messages. Discussion paper no. 106, Institute of Decision Theory and Operation Research, University of Karlsruhe.

Gaul, W. and M. Schader, 1987. Clusterwise aggregation of relations. Discussion Papers in Statistics and Quantitative Economics No. 30, Universität der Bundeswehr, Hamburg (to appear in Applied Stochastic Models and Data Analysis 4, forthcoming).

Gopalan, T.K., 1986. Correspondence analysis and AHC: An exposition of two data analytic techniques. Canadian Journal of Marketing Research 5, 2–9.

Greenacre, M.J., 1984. Theory and applications of correspondence analysis. London: Academic Press.

Guilford, J.P., 1954. Psychometric methods. New York: McGraw-Hill.

Guttman, L., 1946. An approach for quantifying paired comparisons and rank orders. Annals of Mathematical Statistics 17, 144–163.

Hill, M.O., 1974. Correspondence analysis: A neglected multivariate method. Applied Statistics 23, 340–354.

Hoffman, D.L. and G.R. Franke, 1986. Correspondence analysis: Graphical representation of categorical data in marketing research. Journal of Marketing Research 23, 213–227.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24, 417–441 and 498–520.

Krier, M.J. and J.E. Jackson, 1984. Correspondence analysis as a tool for survey statisticians. Toronto: ASA Proceedings of Survey Research Methods Section, 212–217.

Lebart, L., A. Morineau and K.M. Warwick, 1984, Multivariate descriptive statistical analysis. New York: Wiley.

Nishisato, S., 1977. Recent developments in scaling and related areas: A bibliographic review, Japanese Journal of Behaviormetrics 4, 74–95 (in Japanese with English abstract).

Nishisato, S., 1978. Optimal scaling of paired comparison and rank order data: An alternative to Guttman's formulation. Psychometrika 43, 263–271.

Nishisato, S., 1978. Multidimensional scaling: A historical sketch and bibliography. Department of MECA technical report. OISE.

Nishisato, S., 1980. Analysis of categorical data: dual scaling and its applications. Toronto: University of Toronto Press.

Nishisato, S., 1984. Forced classification: A simple application of a quantification method. Psychometrika 49, 25–36.

Nishisato, S., 1986. Quantification of categorical data: A bibliography 1975–1986. Toronto: Micro Stats.

Nishisato, S., 1987. Forced classification procedure of dual scaling: Its mathematical properties. In: Bock, H.H. (ed.), Proceedings of the First Conference of the International Federation of Classification Societies. Aachen: North-Holland, forthcoming.

Nishisato, S. and P.S. Arri, 1975. Nonlinear programming approach to optimal scaling of partially ordered categories. Psychometrika 40, 525–558.

Nishisato, S. and I. Nishisato, 1984. An introduction to dual scaling. Toronto: Micro Stats.

Nishisato, S. and W. Sheu, 1984. A note on dual scaling of successive categories data. Psychometrika 49, 493–500.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine and Journal of Science, Series 6, no. 2, 559–572.

Slater, P., 1960. The analysis of personal preferences. British Journal of Statistical Psychology 3, 119–135.

Tanaka, Y. and C. Kodake, 1980. Computational aspects of optimal scaling for ordered categories. Behaviormetrika 7, 35–47.

Tenenhaus, M. and F.W. Young, 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. Psychometrika 50, 91–119.

Thurstone, L.L., 1927. A law of comparative judgement. Psychological Review 34, 273–286.

Thurstone, L.L., 1931. Rank order as a psychological method. Journal of Experimental Psychology 14, 187–201.

Tucker, L.R., 1960. Intra-individual and inter-individual multidimensionality. In: Guliksen, H. and S. Mersick (eds.), Psychological scaling, New York: Wiley.

Vasserot, G., 1976. L'analyse des correspondances appliquée an marketing: Choix d'un nom d'engrais. Les Caliers de l'Analyse des Données 1, 319–333.

Weingarden, P. and S. Nishisato, 1986. Can a method of rank ordering reproduce paired comparisons? An analysis by dual scaling (correspondence analysis). Canadian Journal of Marketing Research 5, 11–18.