PSYCHOMETRIKA—VOL. 45, NO. 2.
JUNE, 1980

THE JOINT ANALYSIS OF DIRECT RATINGS, PAIRWISE PREFERENCES, AND DISSIMILARITIES

J. O. RAMSAY

MCGILL UNIVERSITY

In studies involving judgments of similarity or dissimilarity, a variety of other variables may also be measured. Examples might be direct ratings of the stimuli, pairwise preference judgments, and physical measurements of the stimuli with respect to various properties. In such cases, there are important advantages to joint analyses of the dissimilarity and collateral variables. A variety of models are described for relating these and algorithms described for fitting these to data. A number of hypothesis tests are developed and an example offered.

Key words: multidimensional scaling, paired comparisons.

1. Introduction

In many experimental situations it is desirable to collect a number of different kinds of information on subject's perceptions and evaluative reactions to a set of stimuli. Judgments of dissimilarity are especially useful for revealing the cognitive aspects of attitudes through multidimensional scaling analysis. On othe other hand, affective or evaluative reactions are usually measured by either direct ratings on "like-dislike" or Likert scales, or by judgments of direction and degree of preference or dominance for one stimulus over another for a set of pairs of stimuli. Judgments of this kind are usually analyzed by a variety of unidimensional scaling procedures to yield scale values for each stimulus. Specific cognitive aspects may also be measured by direct ratings or pairwise preference or dominance. In addition to all this, the experimenter may also wish to incorporate various physical measurements of the stimuli into the analysis.

A central problem in such multivariate experimental designs is to relate the dissimilarity data to the direct ratings, pairwise preferences, or physical measurements. There is usually some reason for supposing that the processes which give rise to the various types of judgments or measurements share features in common. In fact, these shared features may be exactly what is being investigated, so that the experimenter is interested in how the cognitions and physical characteristics of the stimuli give rise to a particular subject's evaluations of these same stimuli. Or, perhaps, he may be interested in how certain physical measurements relate to certain cognitions, with a view to replacing subjective scaling results by "hard" measurements in future investigations.

Here is an experimental situation that illustrates these problems nicely. The Pulp and Paper Research Institute of Canada was concerned about the subjective impact of photographs printed on various types of newsprint by various processes. Some types of reproduction yield pleasing results but are expensive, while others are not as attractive but very desirable for economic or ecological reasons. After some speculation as to the aspects of

The research reported here was supported by grant number APA 320 to the author by the National Research Council of Canada.

Requests for reprints should be sent to James O. Ramsay, Department of Psychology, Stewart Biological Sciences Building, 1205 McGregor Avenue, Montreal, Quebec, CANADA H3A 1B1.

0033-3123/80/0600-2850$00.75/0

newsprint photographs which give rise to these evaluative reactions on the part of various types of readers, a study was launched in which were collected (a) a number of physical measures such as degree of contrast, mottle, "see-through", etc., (b) judgments of dissimilarity among eight sample reproductions of a test photograph, and (c) judgments of the degree to which one reproduction was preferred to another for each of the 28 possible pairs. The goal was to study the aspects of the samples which led to the preferences expressed. The dissimilarity data were included to indicate possible aspects not measured physically and to indicate which physical aspects were noticed by the subjects.

Other studies which are similar to this in design are O'Hare [1976] and Ramsay and Case [1970], and Duder and Aronson [1978].

There is some literature on techniques for relating these types of data. Bechtel [1976] provides a comprehensive overview of previous work and some new contributions. Carroll [1972] deals specifically with the problem of relating direct ratings to the configuration resulting from a multidimensional scaling analysis. Pairwise preferences and direct ratings were related by a combination of multiple regression and factor analysis in Ramsay and Case [1970].

Two models have been investigated in any detail for the relation between direct ratings and the configuration of points produced by multidimensional scaling. These are the *scalar product* or linear model and the *ideal point* or unfolding model. In the former a preferred direction is defined in the configuration and it is postulated that the further along in this direction a point is, the more highly the corresponding stimulus will be rated. In the latter, a preferred location is defined, and the closer a point is to this ideal point, the more highly the stimulus will be rated. If the preferred location is well beyond the configuration of points, the two models become indistinguishable.

For either model, the estimation of the parameters determining the ideal direction or point has usually been conditional on knowing the locations of the points in the configuration, perhaps as a consequence of a previous multidimensional scaling. A principal focus of this paper will be the *joint* estimation of the positions of these points and the ideal direction or point. As will be indicated formally below, the main reason for this approach is the superior quality of the resulting estimates. Joint estimation will in general produce parameter estimates having less bias and sampling variance because all of the available information in the sample is used in their estimation. However, techniques for conditional estimation will also be developed.

The maximum likelihood approach to multidimensional scaling of Ramsay [1977;1978a] will be extended to this situation. The notation to be employed is as follows: n stimuli are to be represented as points in a euclidean space of dimensionality k. The location of the i^{th} point is defined by the coordinates x_{im}, $m = i, \cdots, k$. Data are collected from N subjects. These data include the rating of the degree of dissimilarity d_{ijr} between stimuli i and j by subject r. Moreover, ratings or measurements on a set of L properties or aspects may be involved. The direct rating of stimulus i on property g by subject r will be indicated by u_{igr}. If this is actually a physical measurement, these values will as a rule be the same for all subjects, and the subscript r can be dropped. For convenience, we may focus on only one property at a time, and the subscript g may then be dropped. The pairwise preference for stimulus i over stimulus j with respect to property g for subject r will be indicated by p_{ijgr}.

Each of these pieces of data will be assumed to have a corresponding errorless or model value, and this will be indicated by an asterisk. For example, d^*_{ijr} is the model value for dissimilarity judgement d_{ijr}.

The model for dissimilarity that will be considered explicitly in this paper will be that referred to by Ramsay [1978a, b] as M2:

$$d^*_{ijr} = v_r\left[\sum_m^k (x_{im} - x_{jm})^2\right]^{p_r/2}. \tag{1}$$

Extension of the following results to other models, such as the individually weighted dimension model M3, are straightforward.

Likewise, it will be assumed that the dissimilarities have independent lognormal distributions about their model values, defined by

$$\log d_{ijr} \sim N(\log d^*_{ijr}, \sigma^2). \tag{2}$$

The extension to nonhomogeneous variances by Ramsay [1978a] can also be incorporated into what follows without difficulty.

2. *Direct Ratings and Dissimilarities*

The relationship between direct ratings and the configuration underlying the dissimilarities has been expressed in a variety of ways. The simplest of these is the scalar product model:

$$u^*_{igr} = \sum_m^k a_{grm}x_{im} + c_{gr}. \tag{3}$$

In this model the multipliers a_{grm} can be interpreted as direction cosines or regression coefficients defining an ideal direction specific to property g and subject r. According to this model, if a_{grm} is positive, the more of aspect m underlying the dissimilarities the stimulus has, the more highly it will be rated with respect to property g by subject r. For example, if property g is "clarity" of a newsprint photograph, and aspect m as revealed through the dissimilarities turns out to be contrast, which for a particular subject is weighted positively, then the higher the contrast the higher the subject will rate the clarity.

From a statistician's point of view, the greatest merit of the model is its simple linear structure. Conditional on knowing the coordinates x_{im}, the estimation of the weights is simply a matter of multiple regression. Unfortunately, while useful in many situations, the scalar product model has some serious drawbacks from a psychologist's point of view. Subjects' likings never increase without limit as the model would suggest, and in some instances will obviously decrease after a certain point. All children like sweet things and the sweeter the better, perhaps without limit. But for most adults there is an optimum amount of sugar in anything beyond which utility begins to fall off.

The ideal point model offsets this problem by postulating an ideal location in the space used to represent the stimuli, with the rating for a particular stimulus being a function of how far that stimulus is from this ideal location. Let this ideal location for property g and subject r be determined by the coordinates y_{grm}, $m = 1, \cdots, k$. The distance from this location to the i^{th} point is given by

$$d^*_{igr} = \left[\sum_m^k (y_{grm} - x_{im})^2\right]^{1/2}. \tag{4}$$

The model used by Carroll [1972] and Bechtel [1976] to relate the errorless rating to distance is

$$u^*_{igr} = -b_{gr}d^*_{igr} + c_{gr}. \tag{5}$$

The coefficients b_{gr} and c_{gr} are necessary because as a rule direct ratings are assumed to be on an interval scale.

The author prefers a slightly different formulation:

$$u^*_{igr} = -b_{gr} \log d^*_{igr} + c_{gr}. \tag{6}$$

The logarithmic transformation takes the ratio scale quantity d^* into an interval scale quantity and thus makes it compatible with a linear transformation to the interval scale direct rating. Moreover, the Carroll-Bechtel model (5) predicts an upper bound on the rating at $d^* = 0$ which is inconsistent with the interval scale notion. Under the revised formulation (6) the regression coefficient b_{gr} is invariant under changes of scale of the dissimilarities which can be useful.

Another relation between direct ratings and distance which has some appeal is

$$u^*_{igr} = b_{gr} \exp(-a_{gr} d^*_{igr}) + c_{gr}. \tag{7}$$

This relation accommodates the fact that most response scales are bounded both below and above, and, moreover, the notion of utility and other subjective dimensions being bounded has some appeal on psychological grounds. Although in this paper we shall confine our attention to relation (6), it must be acknowledged that both (5) and (7) as well as other possible relations will seem more reasonable in some situations. The modifications in what follows to accommodate these alternatives are comparatively minor.

These relations imply that the quantities u^*_{igr} depend on the coordinates x_{im} via the quantities d^*_{igr}. Thus, the direct ratings contain at least some information about these coordinates in addition to that offered by the dissimilarity data. From the standpoint of obtaining the best possible estimates of the configuration, therefore, it is important to carry out a joint analysis of the direct ratings and dissimilarities.

The choice of an error distribution for direct ratings of utility involves a variety of considerations. The lognormal assumption (2) for dissimilarities was motivated by the fact that dissimilarities are defined naturally on the positive real line, and (2) can be thought of as arising from a transformation of the response to an interval scale through the use of logarithms, and then assuming normality. Since utility is in principle already on an interval scale, this line of argument would suggest

$$u_{igr} \sim N(u^*_{igr}, \xi^2_g). \tag{8}$$

This assumption may not be too bad where plenty of room is allowed on the response continuum for relatively extreme responses. However, both (2) and (8) will certainly break down where categorical rating scales with a very limited number of categories are employed, or in situations where judgments are likely to be extreme rather frequently. The appropriate assumption in such cases would have to take into account the restriction on the range of responses and their discretization. Takane [1978, Note 2] has taken some important steps toward modelling such situations for dissimilarity data.

The expression of the joint log likelihood for the two sets of data employing (2) and (8) can be simplified considerably by defining the following two error sums of squares:

$$Q_r = \sum_{i \neq j}^{n\,n} (\log d_{ijr} - \log d^*_{ijr})^2, \tag{9}$$

and

$$S_{gr} = \sum_{i}^{n} (u_{igr} - u^*_{igr})^2. \tag{10}$$

The log likelihood for the dissimilarity data then becomes

$$\log L_d = -M_d \log\sigma - \frac{1}{2}\sigma^{-2} \sum_r^N Q_r, \tag{11}$$

where M_d is the total number of dissimilarity judgments; and the log likelihood for the utility data is

$$\log L_u = -\sum_g^L \left[M_g \log \xi_g + \frac{1}{2}\xi_g^{-2} \sum_r^N S_{gr} \right], \tag{12}$$

where M_g is the total number of direct ratings for the g^{th} property. The summations in (9) to (12) can be understood to be only over those observations which have been made, so that missing data presents no particular problems. If the assumption that the errors for the utility judgments are independent of those for the dissimilarities, then the joint likelihood is given simply by

$$\log L = \log L_d + \log L_u. \tag{13}$$

Maximum likelihood estimates for the dispersion parameters are as follows:

$$\sigma^2 = M_d^{-1} \sum_r^N Q_r, \tag{14}$$

and

$$\xi_g^2 = M_g^{-1} \sum_r^N S_{gr}. \tag{15}$$

Thus, their estimation conditional on the coordinates x_{im} and the regression coefficients a_{grm}, b_{gr}, and c_{gr} is a simple matter. Likewise, the regression coefficients can be estimated by maximum likelihood through the minimizing of the error sum of squares S_{gr}. Since these coefficients enter linearly into u^*_{igr}, this is a conventional linear least squares problem. The estimation of the dissimilarity parameters log v_r and p_r is also a linear least squares problem since they enter linearly into log d^*_{ijr}.

Consequently, the troublesome parameters in this data analysis problem are the coordinates of the ideal points y_{grm}. An iterative approach must be adopted here, with a strategy which updates the configuration and ideal point coordinates each iteration and then estimates the remaining parameters conditional on these.

There are many ways to tackle the problem of estimating the configuration and the ideal points. Ramsay [1977, 1978b] adopted an implicit equation approach which had a certain relation to the gradient method. This approach has also been called the C-matrix method by some authors. It has the advantage of requiring only a moderate amount of computer memory, which is an important consideration in view of the very large number of parameters that may be involved in some problems. Ramsay [1980] suggests that this method converges reliably and reasonably quickly for model M2 in multidimensional scaling.

There are, however, some important advantages to using some variation of a Newton-Raphson approach. First of all, an estimate of the variance-covariance matrix for some or all of the parameter estimates is very conveniently available as a byproduct of such iterations. Secondly, it is easier to modify the algorithm to permit equality or inequality constraints on the parameters than in the implicit equation approach. Thirdly, such approaches converge much more rapidly when very near the solution to the likelihood equations. This virtue may be questionable, since parameter estimates need be accurate in most cases only to about two or three significant digits, and the log likelihood is required to no more than one decimal place of accuracy when used to compute chi-square

statistics. Also, although far fewer Newton-Raphson iterations are typically required than C-matrix iterations, the overhead per iteration can more than offset this saving, and the author's experience indicates that there is very seldom any saving in computation time by using Newton-Raphson iterations.

In the case of the ideal point model, there are two sets of parameters which require iterative estimation. The approach used by the author has been to update the configuration coordinates x_{im} by either C-matrix or Newton-Raphson iterations, and then to iterate to convergence or near convergence for the ideal point coordinate y_{grm}.

In the case of the ideal point model, there are two sets of parameters which require iterative estimation. The approach used by the author has been, for each global or master iteration, to proceed through a fixed number of updates of the configuration coordinates x_{im}, holding the ideal point coordinates y_{grm} and all other parameters fixed, and then to proceed through a fixed number of ideal point coordinate updates holding the configuration coordinates fixed. This can be called an alternating least squares approach. Each master iteration terminates with the estimation of the regression coefficients and dispersion parameters conditional on the configuration and ideal point estimates.

The maximum likelihood estimation problem for the configuration and ideal points can be reduced in this and the following section to nonlinear least squares problems. If Newton-Raphson iterations are to be used, it has turned out to be useful to use the expected value of the second derivative matrix or Hessian. This both guarantees positive semidefiniteness and reduces the computation load since such a matrix does not depend on the data. The use of this matrix to define a direction of search seems to be about as effective as that defined by the Hessian itself which, it can be shown, need not be positive semidefinite. Finally, the Moore-Penrose inverse of this matrix is an estimate of the asymptotic variance-covariance matrix of the parameter estimates [Ramsay, 1978a]. The use of the expected second derivative matrix is equivalent to the Gauss-Newton procedure for solving nonlinear least squares problems and, in this instance, to Fisher's scoring method for maximum likelihood estimation.

The estimation of the ideal point coordinates requires the minimizing of the error sum of square S_{gr} defined in (10), and in the case of the relation (6), the first derivatives are given by

$$\frac{\partial S_{gr}}{\partial y_{grq}} = -2 \sum_{i}^{n} (u_{igr} - u^*_{igr}) \frac{\partial u^*_{igr}}{\partial y_{grq}}, \tag{16}$$

where

$$\frac{\partial u^*_{igr}}{\partial y_{grq}} = -b_{gr} (y_{grq} - x_{iq}) d^{*-2}_{igr}. \tag{17}$$

The expected second derivative matrix is given by

$$E \frac{\partial^2 S_{gr}}{\partial y_{grq} \partial y_{grt}} = 2 \sum_{i}^{n} \frac{\partial u^*_{igr}}{\partial y_{grq}} \frac{\partial u^*_{igr}}{\partial y_{grt}} \tag{18}$$

Thus, the ideal point for each subject and each directly rated property can be estimated independently of the other conditional on the configuration coordinates x_{im}. In the author's experience these iterations converge rapidly and reliably unless the starting values for these iterations are very unsatisfactory.

The following procedure seems to yield quite satisfactory starting estimates of the ideal points. First, compute the parameters a_{grm} in the scalar product model by solving the linear least squares problem implied by minimizing S_{gr} with respect to the a_{grm}'s. Then compute a point in this preferred direction by moving away from the origin in this direc-

tion by a judicious amount. Such a point can be produced by using the equation

$$y_{grq} = a_{grq} \max_i \left\{ \left(\sum_m x_{im}^2 \right)^{1/2} \right\} \tag{19}$$

assuming that the origin is at the centroid of the configuration. Note that when the true location of the ideal point is close to the origin, the scalar product model will be grossly inappropriate. As a consequence, the regression coefficients a_{grm} will be all near zero and the estimated ideal point will be near the origin as desired.

We turn now to the problem of estimating the configuration coordinates x_{im}. Note that both of the terms of the log likelihood (13) depend on these coordinates via the two sets of distances d^*_{igr} and d^*_{ijr} or, in the scalar product model, via relation (3). Here a particular problem arises; the invariance of the log likelihood with respect to translation and reflection of the coordinates. This arises because it depends on the coordinates only via these two types of distance, or via distance within the configuration and relation (3), and these quantities are invariant under such transformations. Thus, the expected second derivative matrix will at best be positive semidefinite, with rank at most $nk - k$. Since the conventional Newton-Raphson approach requires that this matrix be positive definite, some modification of this approach is in order.

The solution to this difficulty is to employ the Moore-Penrose inverse of the expected second derivative matrix. This provides an update to the parameter estimates which, when close to the solution, provides quadratic convergence to a solution, and at the same time provides an update to the current estimate which as minimum length and is therefore useful from the standpoint of testing for satisfactory convergence. Moreover, as Ramsay [1978a] has shown, the Moore-Penrose inverse provides in a certain sense the asymptotic variance-covariance matrix of the estimates. An algorithm for computing the Moore-Penrose inverse of a symmetric positive semidefinite matrix is described in the appendix.

The maximum likelihood estimation of the coordinates x_{pq} then reduces to the minimization of the following weighted error sum of squares:

$$T = \sigma^{-2} \sum_r Q_r + \sum \xi_g^{-2} \sum_r S_{gr}. \tag{20}$$

The first and the expected second derivatives of this quantity depend on the first derivatives of $\log d^*_{ijr}$ and u^*_{igr}. For the former quantity this is

$$\frac{\partial \log d^*_{pjr}}{\partial x_{pq}} = \frac{\partial \log d^*_{jpr}}{\partial x_{pq}} = \frac{p_r(x_{pq} - x_{jq})}{\sum_m (x_{pm} - x_{jm})^2} \tag{21}$$

In the case of the scalar product model the first derivative of u^*_{igr} is

$$\frac{\partial u^*_{igr}}{\partial x_{pq}} = \begin{cases} a_{grq}, & i = p. \\ 0, & i \neq p. \end{cases} \tag{22}$$

In the case of the ideal point model,

$$\frac{\partial u^*_{igr}}{\partial x_{pq}} = \begin{cases} b_{gr}(y_{grq} - x_{pq})\, d^{*-2}_{pgr}, & i = p, \\ 0, & i \neq p. \end{cases} \tag{23}$$

It is worth noting that (20) displays the maximum likelihood estimation process as a weighted least squares problem in which the variances appear explicitly. Thus, we cannot reduce the objective function to one which does not involve the variance parameters in

the way that Ramsay [1977] did, and these must be estimated simultaneously with the other parameters in the analysis. This is characteristic of weighted least squares problems in which weights must be estimated.

If one substitutes (14) and (15) into (13), then the final value of the joint log likelihood reduces to

$$\log L = -\tfrac{1}{2}M_d\left(1 + \log M_d^{-1}\sum_r Q_r\right) - \tfrac{1}{2}\sum_g M_g\left(1 + \log M_g^{-1}\sum_r S_{gr}\right). \tag{24}$$

This form is useful for constructing various test statistics to be described in the next section.

3. Testing Various Hypotheses

One of the main advantages of maximum likelihood estimation is the possibility of large-sample tests of restricted models against those of which they are special cases. In the case of the models described in this paper, one must proceed with some caution since subject-specific parameters are used for modelling both utilities and dissimilarities, implying that new parameters are added for each subject as well as with each new stimulus or property being rated. This makes a truly asymptotic sample size impossible since the number of parameters cannot be a fixed quantity. Ramsay [1979] has collected Monte Carlo evidence to show that applying large sample tests in these situations leads to some bias in favor of rejections of the null hypothesis. Fortunately, there seems to be some hope that the correction for this bias is fairly simple, and these tests can be used profitably if one is suitably careful.

To test whether the fit in k dimensions is superior to that in $k-1$ dimensions, twice the difference between log likelihoods can be assessed against a chi squared critical value. To compute the appropriate number of degrees of freedom, one must take the difference between the number of mathematically independent parameters being fit in each case. For k dimensions, there are nk values of x subject to rotational and translational invariance. Thus there are $nk - k(k+1)/2$ parameters here. There are also N v_r's and N p_r's. However, the constraint $\sum \log v_r = 0$ is imposed in order to remove a scale trade-off between the v's and the x's. Finally, there are $LN(k+1)$ regression coefficients to compute in the case of the scalar product model, and $LN(k+2)$ ideal point coordinates and regression coefficients for the ideal point model. Thus, for a k-dimensional fit there are a total of $nk - k(k+1)/2 + 2N - 1 + LN(k+1)$ and $nk - k(k+1)/2 + 2N - 1 + LN(k+2)$ parameters for the scalar product and ideal point models, respectively, excluding variance parameters. The difference between the number of parameters being estimated for fits in k and $k-1$ dimensions works out to $n - k + LN$ in either case.

To test the null hypothesis that the direct ratings are unrelated to the dissimilarity judgments, it is necessary to carry out a multidimensional scaling analysis of the dissimilarity data alone. This is equivalent to minimizing the quantity $\sum Q_r$. Let the value Q_r obtained by this minimizing with respect to the dissimilarities alone indicated by Q'_r, while the value Q_r obtained by the joint analysis of the direct ratings and dissimilarities will be indicated simply by Q_r. Similarly, let the value of the quantities S_{gr} obtained by using $u^*_{igr} = u_{.gr}$, the mean direct ratings averaged over stimuli, be indicated by S'_{gr}. Then the statistic which has an asymptotic chi square distribution and which tests the null hypothesis that $a_{grm} = 0$, $m = 1, \cdots, k$, in the case of the scalar product model, or $b_{gr} = 0$ in the case of the ideal point model, is given by

$$\chi^2 = \sum_g M_g\left(\log \sum_r S'_{gr} - \log \sum_r S_{gr}\right) + M_d\left(\log \sum_r Q'_r - \log \sum_r Q_r\right). \tag{25}$$

The degrees of freedom for the scalar product model are LNk while for the ideal point model they are $LN(k + 1)$.

Another hypothesis that is of potential interest is whether or not the ideal point model provides a superior fit to that of the scalar product model. This also can be tested by computing a chi square statistic since the former reduces to the latter when the ideal point is far away from the configuration. The ideal point has LN more parameters than the scalar product model, so that the doubled difference between log likelihoods can be assessed against the tabled chi square critical value with this many degrees of freedom.

Many investigators will want to carry out these hypothesis tests for each direct rating in turn. In such cases, it will be necessary to set $L = 1$, and carry out a joint analysis for each rating separately. However, since direct ratings are usually highly correlated, one should remember that such hypothesis tests are not independent of each other.

4. Pairwise Preferences and Dissimilarities

There are some sound reasons for collecting pairwise preference or dominance judgments with respect to some property rather than ratings with respect to the property directly. In such judgments, the subject is presented with stimuli i and j, and asked to rate the degree to which he prefers i to j, or the degree to which i has more of the property than j. If, in fact, stimulus j dominates, this rating is given a negative number. Let this rating for subject r and property g be indicated by p_{ijgr}.

One of the advantages of pairwise ratings is that they can be reduced to scale values for each stimulus and, in the process, an internal consistency index can be computed which has a variety of uses. The usual model for relating preferences to direct ratings is

$$p^*_{ijgq} = u^*_{igr} - u^*_{jgr} . \tag{26}$$

This is a linear model if the u's are to be estimated, and a multiple correlation coefficient can be computed to indicate when subjects are having difficulty with the rating task. An F-ratio can be computed to assess inter-group differences and other effects [Bechtel, 1976].

Because there are many more possible pairwise preference judgments than direct ratings, the former offer the advantages of collecting more data for estimation of parameters of interest. Finally, subjects often find preference judgments easier to make because the elements to be compared are explicitly presented, whereas with direct ratings a stimulus has to be compared with an implicit standard which can change from time to time or subject to subject.

The extension of the scalar product model (3) to preferences yields

$$p^*_{ijr} = \sum_{m}^{k} a_{grm}(x_{im} - x_{jm}) , \tag{27}$$

while the ideal point model for preferences becomes

$$p^*_{ijgr} = b_{gr} \log \left(\frac{d^*_{igr}}{d^*_{jgr}} \right) . \tag{28}$$

With the distributional assumption

$$p_{ijgr} \sim N\left(p^*_{ijgr} , \zeta^2_g \right) , \tag{29}$$

the log likelihood for the data including both preferences and dissimilarities becomes

$$\log L = -M_d \log \sigma - \tfrac{1}{2}\sigma^{-2} \sum_{r}^{N} Q_r - \sum_{g}^{L} M_g \log \zeta_g - \tfrac{1}{2} \sum_{r}^{L} S_{gr} \tag{30}$$

where M_g is now the total number of preference observations, and S_{gr} is now

$$S_{gr} = \sum_{i \neq j}^{n} \sum^{n} (p_{ijgr} - p^*_{ijgr})^2 . \tag{31}$$

As with direct ratings, these are linear models given the configuration coordinates x_{im} and ideal point coordinates y_{grm}, and the remaining parameters can be estimated in the usual way conditional on these.

The first derivatives of the quantities S_{gr} which must be minimized with respect to ideal point coordinates are

$$\frac{\partial S_{gr}}{\partial y_{grq}} = -2 \sum_{i \neq j}^{n} \sum^{n} (p_{ijgr} - p^*_{ijgr}) \frac{\partial p^*_{igr}}{\partial y_{grq}} , \tag{32}$$

where

$$\frac{\partial p_{ijgr}}{\partial y_{grq}} = -b_{gr}[(y_{grq} - x_{iq}) d^{*-2}_{igr} - (y_{grq} - x_{jq}) d^{*-2}_{jgr}]. \tag{33}$$

The expected second derivative matrix is

$$E \frac{\partial^2 S_{gr}}{\partial y_{grq} \partial y_{grt}} = 2 \sum_{i \neq j}^{n} \sum^{n} \frac{\partial p^*_{ijgr}}{\partial y_{grq} \partial y_{grt}} . \tag{34}$$

The discussion of techniques for optimization and starting values for the ideal point coordinates applies here as well.

The estimation of configuration coordinates x_{im} requires the minimization of the quantity T defined in (20). The first derivatives of p^*_{pjgr} for the scalar product and the ideal point model are given by the right sides of (22) and (23) respectively, and in both cases.

$$\frac{\partial p^*_{jpqr}}{\partial x_{pq}} = - \frac{\partial p^*_{pjgr}}{\partial x_{pq}} \tag{35}$$

The remarks in Section 3 concerning hypothesis testing apply to the joint analysis of preferences and dissimilarities as well. An additional benefit from obtaining preferences as opposed to direct ratings is that one can fit the model (26) to the preferences by a separate analysis. This permits the comparison between the joint model and the model in which (27) or (28) is fit to the preferences and Model M2 is fit to the dissimilarities. This comparison permits us to see whether there is additional variability in the preferences over and above that fit by the joint model. The chi square statistic involved has $LN(n - k - 1)$ degrees of freedom for the scalar product model and $LN(n - k - 2)$ for the unfolding model.

5. Physical Measurements and Dissimilarities

Physical measurements on a set of stimuli can be related to dissimilarity judgments by viewing them as direct ratings which are identical for each subject, and employing the methods of Section 2. However, it may also be desirable to view them as coordinates for the stimuli and thus as defining locations of a set of points with respect to some coordinate system. From this point of view, the dissimilarities are to be related to the interpoint distances among these points.

When physical measurements are viewed as coordinates, there is usually no particular reason to suppose that the axes of the space are orthogonal. Indeed, there is usually good reason to suppose that a particular pair of measurements are subjectively related.

Thus, each subject's metric must be estimated in the following relation between dissimilarity and generalized distance:

$$d_{ijr} \approx d^*_{ijr} = v_r \left[\sum_p \sum_s (x_{ip} - x_{jp}) w_{rps} (x_{is} - x_{js}) \right]^{p_r/2}. \tag{36}$$

The matrix W_r defining the metric will be positive semidefinite, and following Bloxom [1978] it is convenient to decompose it into

$$W_r = Z_r Z_r^t \tag{37}$$

where Z_r is lower triangular. The problem of fitting the dissimilarity observations then becomes one of estimating the $k(k + 1)/2$ elements of Z_r as well as the regression coefficients v_r and exponent p_r.

Under the hypothesis of lognormal distribution of d_{ijr} about d^*_{ijr}, the maximum likelihood estimation of Z_r requires the minimization of the function Q_r as defined in (9). In this instance, the derivative of Q_r with respect to z_{pqr}, $p \geq q$, is

$$\frac{\partial Q_r}{\partial z_{pqr}} = 2p_r \sum\sum_{i \neq j} (\log d_{ijr} - \log d^*_{ijr}) d^{*-2}_{ijr} (x_{ip} - x_{jp}) \sum_{m \geq q} (x_{im} - x_{jm}) z_{mqr}, \tag{38}$$

for $p \geq q$, and the expected second derivative matrix is

$$E \frac{\partial^2 Q_r}{\partial z_{pqr} \partial z_{str}} = 2p_r \sum\sum_{i \neq j} d^{*-4}_{ijr} (x_{is} - x_{js})(x_{ip} - x_{jp}) \times \left[\sum_{m \geq q} (x_{im} - x_{jm}) z_{mqr} \right] \left[\sum_{m \geq t} (x_{im} - x_{km}) z_{mtr} \right], \tag{39}$$

for $p \geq q$ and $s \geq t$.

A variety of useful tests can be performed to compare the results of this analysis with various others. For example, the fit using k physical measurements can be compared with that using $k - 1$ measurements, and the resulting chi square statistic will have Nk degrees of freedom. The result of a multidimensional scaling analysis using Model M2 can also be compared with this analysis for a single subject as a test of whether the physical measurements used account for all the systematic variation in the dissimilarities. The resulting chi square would have $nk - 1$ degrees of freedom. It must be borne in mind, however, that this would be a small sample test and the tabled chi square criterion should probably be increased somewhat.

6. *An Example: The Perceptions of Newsprint Photographs*

In this example, 14 subjects who were technicians in newsprint manufacturing rated all possible pairs of a single photograph reproduced on eight newsprint samples. Each pair was rated both in terms of the degree of perceived dissimilarity (on a 25 category rating scale) and in terms of degree of preference from the point of view of general acceptability. In addition, measurements of the amount of mottle, show-through, and contrast were taken on each sample by photometric techniques.

Various log likelihoods for these data are displayed in Table 1, along with estimates of the standard errors for dissimilarity and preferences. A comparison of the results for the analysis of the dissimilarities along in two and three dimensions produces a chi square of 2.4 with 5 degrees of freedom, and this indicates that two dimensions are sufficient to account for the dissimilarity data. However, in comparing the two- and three-dimensional solutions for the joint analysis of preference and dissimilarity data using the scalar product model, the chi square statistic is 47.8 with 19 degrees of freedom, which is significant

TABLE 1

Some Statistics for the Print Quality Data

Type of Analysis	Dimensions	Std. Error for Diss.	Std. Error for Pref.	Log Likelihood
Multidimensional Scaling (M2)	2	.566*	-	48.4
Multidimensional Scaling (M2)	3	.568*	-	49.6
Joint Analysis (Unfolding)	2	.544	3.22	-611.8
Joint Analysis (Unfolding)	3	.543	3.01	-585.4
Joint Analysis (Scalar Product)	2	.543	3.34	-625.5
Joint Analysis (Scalar Product)	3	.539	3.17	-601.6
Preference Analysis	-	-	3.50*	-567.2

*These standard error estimates are unbiased.

at the .01 level and indicates the need for a three- or higher dimensional solution. Results for the unfolding model analysis are very similar. This could indicate that an additional property determined the preference data over and above those which were the basis for the dissimilarity judgments.

A comparison of the unfolding model log likelihood with the scalar product model log likelihood in three dimensions produces a chi square of 32.4 with 14 degrees of freedom, and this is significant at the .01 level. This provides some evidence that the unfolding model is more appropriate for these data. An additional question is whether or not the separate analyses of the two sets of data produces a superior fit to that produced by the joint analysis. Using the unfolding joint analysis in three dimensions and the M2 analysis in two dimensions, the resulting chi square has a value $2[585.4 - (-567.2 + 48.4)] = 133.2$ with 37 degrees of freedom. This is very significant, and indicates that there is variability in the preference data not accommodated by the joint unfolding model. Nevertheless, an examination of the standard error estimates for the preference judgments which are displayed in Table 1 indicates that in practical terms the degree of fit of the unfolding model is quite comparable to that obtained by a separate analysis of the preference data.

As a final note on these data, it is interesting to observe that when the preferences were fit by the model

$$\text{sgn}(p_{ijr})|p_{ijr}|^{q_r} \approx p^*_{ijr} = u^*_{ir} - u^*_{jr}, \tag{40}$$

where the exponent q_r for the preferences was estimated by maximum likelihood along with the scale values u_{ir}, the average exponent was 1.38 with a standard deviation of 0.26. This indicates a deficiency in model (26) which fails to allow for a nonlinear relation be-

tween data and model. It also seems probable that the models for direct ratings should also be extended in this way.

The dissimilarity ratings were also related to the measurements of mottle, show-through, and contrast through model (36). The results are shown in Table 2. Here it can be seen from the eigenvalues of W that all subjects employed only two dimensions. However, the amount of weight attached to the physical measurements varied greatly from subject to subject. For example, Table 3 displays the metrics W for Subjects 2 and 3, and it can be seen that show-through is de-emphasized by the former while contrast is ignored

TABLE 2

Results for Fitting Dissimilarity Data with

Physical coordinates and Individualized Metrics

Subject	Error	Eigenvalues of W d_1	d_2	d_3	Std. Error M2(3)
1	.28	5.06	1.94	.00	.26
2	.32	5.65	2.09	.00	.45
3	.33	5.98	.84	.13	.46
4	.44	4.56	1.33	.00	.58
5	.44	9.00	1.18	.02	.53
6.	.46	3.86	3.63	.00	.59
7	.46	3.38	1.53	.03	.45
8.	.46	7.36	.60	.00	.66
9	.47	2.88	1.83	.00	.63
10	.48	4.14	1.24	.14	.55
11	.48	3.00	1.18	.00	.60
12	.52	2.72	1.33	.00	.67
13	.55	5.48	.76	.00	.69
14	.60	3.75	1.48	.00	.66
Means	.45	4.70	1.50	.02	.56

Table 3

Metric Matrix W for Two Subjects Using Physical Coordinates

Subject	Measurement	Metric Matrix			Correlation Matrix		
2	Mottle	4.2	.8	-1.6	1.00	.41	-.48
	Show-through	.8	.9	-1.5	.41	1.00	-.98
	Contrast	-1.6	-1.5	2.6	-.48	-.98	1.00
3	Mottle	5.3	-1.7	.2	1.00	-.67	.14
	Show-through	-1.7	1.2	-.4	-.67	1.00	-.58
	Contrast	.2	-.4	.4	.14	-.58	1.00

by the latter. Moreover, when these metrics are expressed as correlations by dividing off-diagonal entries by the product of square roots of corresponding diagonal entries, it can be seen that both subjects see the physical dimensions as being nonorthogonal, with the perceived relationship between show-through and contrast being especially strong for Subject 2. A comparison in Table 2 of the unbiased standard error estimates with those obtained from the M2 analysis in three dimensions shows that using the physical measurements along with individualized metric results in a substantially better fit to the dissimilarity data for most subjects.

7. *Extensions of the Ideal Point Model*

The ideal point model permits the fitting of two classes of points: those whose locations are the same for all subjects, and those whose locations are estimated separately for each subject. Thus, it can be viewed as a particular application of a multi-dimensional scaling problem in which the domain of estimation is not the same for each point. One might wish to estimate some points uniquely for each subject and others uniquely for each of a number of subgroups of subjects, while the locations of some points are common to all subjects. Moreover, the use of direct ratings or preferences to fix ideal points is only one way in which idiosyncratic points can be determined. For example, Taylor, Bassili and Aboud [1973] employed "myself" as a stimulus in a study of ethnic group perceptions. Although this point was given a common location, this study invites individualized locations for "myself". It is to be hoped that future software developments will permit more flexibility in choosing the domain of estimation for points.

The use of ideal points can also expand the range of interesting individual differences models for dissimilarity. For example, it is reasonable to suppose that dissimilarities among stimuli "close" to a particular subject's ideal point are perceived differently from those perceptually far removed [Bookstein, Note 3]. For example, one might argue that dissimilarities among fields of science similar to one's own are magnified, whereas those among fields remote from one's interest are diminished. To capture this idea algebraically in two dimensions, let the coordinates for point i and subject r be (z_{ir1}, z_{ir2}) and the corresponding coordinates for the common space be (x_{i1}, x_{i2}). The two sets of coordinates are related as follows:

$$z_{ir1} = t_{ir} \cos \theta_{ir},$$

and

$$z_{ir2} = t_{ir} \sin \theta_{ir},$$

where

$$\theta_{ir} = \tan^{-1} \frac{x_{i2} - y_{r2}}{x_{i1} - y_{r1}}, \; x_{i1} - y_{r1} > 0,$$

$$= 180° + \tan^{-1} \frac{x_{i2} - y_{r2}}{x_{i1} - y_{r1}}, \; x_{i1} - y_{r1} < 0,$$

and

$$t_{ir} = h_r \left(\frac{d^*_{ir}}{h_r} \right)^{w_r}, \tag{41}$$

where d^*_{ir} is the distance from point i to subject r's ideal point in terms of the group coordinates. Figure 1 shows the geometrical interpretation of such a model. In the case that the exponent w_r is less than one, the idiosyncratic location of a point for any subject is moved toward the ideal point if it is more than h_r units away and moved away from the ideal point if it is closer than h_r units. The size of the movement is determined by the exponent w_r; from the figure it can be seen that the distance between Points 1 and 2 is larger for this particular subject than what it is in the common space, while the distance between Points

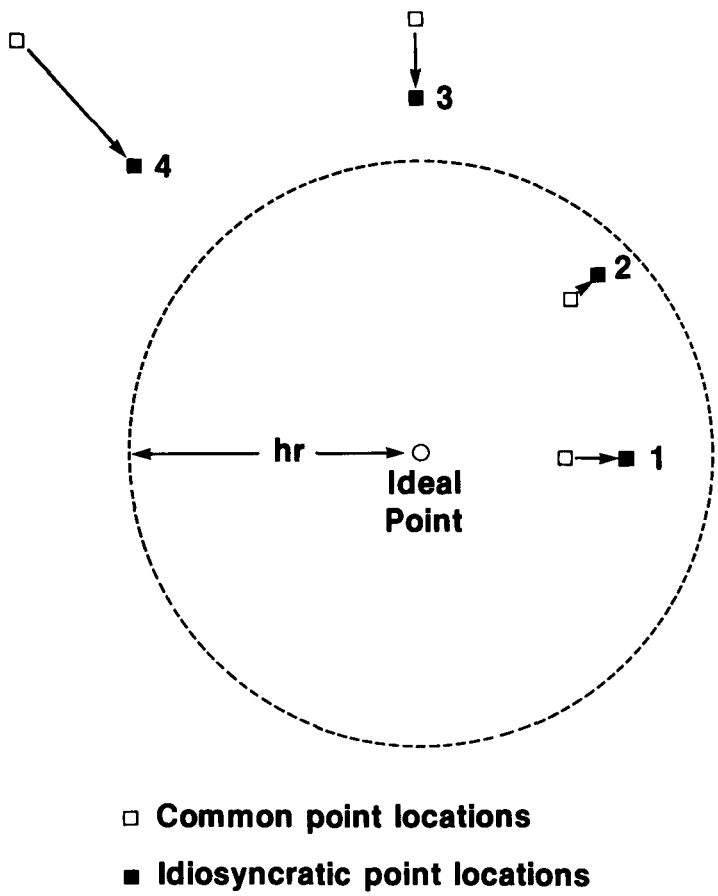

FIGURE 1
The relations between common and idiosyncratic point locations for a particular subject ($w_r = 0.5$).

3 and 4 is relatively smaller. The model can, of course, be generalized to employ other functions besides power transformations. The main advantage of expressing the transformation in polar coordinate form is that this preserves the rotational invariance of the original model and thus makes them comparable.

By involving the ideal point in the model for dissimilarities as well as for direct ratings or preferences, joint analysis of the various types of data is rendered necessary as well as desirable. It may be that in the future multidimensional scaling studies will routinely involve the collection of preferences or direct ratings in addition to dissimilarities in order to locate ideal points as well as a common configuration.

Appendix

An algorithm for computing the Moore-Penrose inverse should be fast, numerically stable, and require as little core as possible. Not all of these objectives can be achieved to optimal degree by any one algorithm, of course, but none of them can be ignored. The application described in this paper can involve some rather large matrices, so that, for example, the well-known procedure of performing an eigenanalysis, inverting positive eigenvalues, and reconstituting the matrix may not be sufficiently fast and may require too much memory.

Let symmetric positive semidefinite matrix P be of order N and rank K. The value of K may be known (as in this application) or computed. The algorithm proceeds by first decomposing P into the product LL', where L is in lower trapezoidal form. That is, L has the structure

$$L = \begin{matrix} T & 0 \\ V & 0 \end{matrix},$$

where T is lower triangular of dimension K and V is $N - K$ by K. This decomposition is performed by Choleski decomposition with pivoting. The rank K can be computed by testing the relative magnitude of the pivotal element.

It can be shown that $P^+ = L(L'L)^{-1}(L'L)^{-1}L'$. This product is expensive to compute in this form, however, and it is more efficient to proceed via the following matrix computations.

(i) replace T by T^{-1}
(ii) replace V by $U = VT^{-1}$
(iii) perform the triangular decomposition $XX' = I + UU'$
(iv) replace X by X^{-1}
(v) compute $W = X^{-1}U$
(vi) compute $R = T^{-1}(I - W'W)$
(vii) compute $Y = RR'$
(viii) compute UY and UYU'.

The Moore-Penrose inverse is then given by:

$$P^+ = \begin{matrix} Y & YU' \\ UY & UYU' \end{matrix}$$

As described above the algorithm requires $N^2 + K(K-1)/2$ words of storage. This can be reduced to N^2 words by computing the elements of W as needed in (vi) which permits the product (vii) to be computed in situ. However, if there is much difference between N and

K the additional computation required can be prohibitive. For very large matrices auxiliary storage can be used in step (vi) which will also make N^2 words suffice.

A listing and documentation of a FORTRAN program to compute P^+ is available on request from the author. If it is only desired to compute the solution P^+B to the linear equation $PX = B$, considerable savings in computer time and memory can be achieved by using the IBM Scientific Subroutine Package subroutines MFSS and MLSS [Note 1].

REFERENCE NOTES

1. *System/360 Scientific Subroutine Package, Version III.* White Plains, N. Y.: International Business Machines Corporation, 1970.
2. Takane, Y. *Multidimensional successive categories scaling: A maximum likelihood method.* Paper presented at the joint meeting of the Psychometric Society and the European Mathematical Psychology Group, Uppsala, Sweden, 1978.
3. Bookstein, A. Personal communication, 1978.

REFERENCES

Bechtel, G. *Multidimensional preference scaling.* The Hague: Mouton, 1976.

Bloxom, B. Constrained multidimensional scaling in N spaces. *Psychometrika,* 1978, *43,* 397–408.

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.). *Multidimensional scaling: Theory and applications in the behavioral sciences, Vol. 1.* New York: Seminar Press, 1972, 105–153.

Duder, S. & Aronson, J. Values in the admission process: An application of multidimensional scaling. *Canadian Journal of Social Work Education,* 1978, *4,* 56–75.

O'Hare, D. Individual differences in perceived similarity and preference for visual art: A multidimensional scaling analysis. *Perception & Psychophysics,* 1976, *20,* 445–452.

Ramsay, J. O. & Case, B. Attitude measurement and the linear model. *Psychological Bulletin,* 1970, *74,* 185–192.

Ramsay, J. O. Maximum likelihood estimation in multidimensional scaling. *Psychometrika,* 1977, *42,* 241–266.

Ramsay, J. O. Confidence regions for multidimensional scaling analysis. *Psychometrika,* 1978, *43,* 145–160 (a).

Ramsay, J. O. *MULTISCALE: Four programs for multidimensional scaling by the method of maximum likelihood.* Chicago: National Educational Resources, Inc., 1978 (b).

Ramsay, J. O. Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika,* 1980, *45,* 141–146.

Taylor, D. M., Bassili, J. N., & Aboud, F. E. Dimensions of ethnic identity: An example from Quebec. *Journal of Social Psychology,* 1973, *89,* 185–192.

Manuscript received 5/7/79
Final version received 12/21/79