4 Interpreting Configurations

'Do you know what this is?'
'No' said Piglet
'It's an A'
'Oh', said Piglet
'Not O—A' said Eeyore severely

A. A. MILNE (The House at Pooh Corner)

The last chapter demonstrated how it is possible to construct a picture of a set of data in the form of a configuration of points, the distances between which represent the dissimilarity between the objects on which the data were collected. Such a representation is easier to assimilate than a matrix of coefficients, but we do not simply want to provide a picture but also, hopefully, to discover hitherto unremarked or unnoticed characteristics of the data. In other words we wish to interpret the configuration.

4.1 What Information is Significant and Stable?

Our interpretation is essentially a two-stage process. First, we look for significant patterns in the configuration, i.e. detect structure, and secondly we ascribe a meaning or interpretation to those patterns or structures. In so doing, it is important that the patterns make use of features of the configuration that are not simply arbitrary artefacts of the scaling procedure. The first questions, then, are: what information is significant and which aspects are stable in a configuration obtained from the basic MDS model? This restriction is important since output from other models sometimes preserves different significant information. The most basic significant information in the solution is the set of relative distances between the points. In the program these distances were calculated from a configuration which has a number of arbitrary, conventional, characteristics. In particular, the configuration was rotated to principal axes (see 3.5.5), and its actual size is also arbitrary, since it is the relative and not the absolute distances which are pertinent. This means that we may legitimately change the axes and the co-ordinate values so long as the relative distances remain unchanged.

The changes or transformations which may legitimately be made to a configuration in the case of the basic model are referred to as a similarity transformation: these consist of rotation, reflection, rescaling and translation (the three 'rs', with a foreign language for good measure?). These turn out to be of vital importance in Chapter 7 when we discuss the comparison of configurations. A detailed definition of similarity transformations is contained in Appendix A7.1 but an intuitive description can be given as follows.

Rotation of axes

MDS solutions using the Euclidean distance model are invariant under rotation—

i.e. the significant distance information is in no way changed by rigid rotation of the axes. *Any* set of axes which are orthogonal (at 90° to each other) will do as well as the axes given in the final configuration from an MDS program.

Reflection of axes
Consider a configuration reproduced on a photographic slide. It does not affect the distances in any way if the slide is looked at from the front or from the back, or indeed whether the slide is the right way up. Technically a reflection occurs when the positive or negative signs of the co-ordinate values on a given dimension are systematically changed.

Rescaling
The co-ordinates of the space may be *uniformly* rescaled, that is stretched or shrunk without destroying the significant information in the solution. This is simply another way of saying that relative distances are ratio-level quantities and may be multiplied by a constant.

Translation of origin
The origin, or zero point on all the axes, may be located at will within the space without changing the distances. (Note however that this does not apply for factor, or vector, solutions).

Clearly we do not want to interpret or make use of information in the configuration which may be affected by any of these transformations, and in the same way we do not want to make use of aspects of the configuration which are not stable.

We noted in Chapter 3 that the position of a particular point in the configuration was not uniquely fixed by the procedure, but rather fixed within a portion of the space known as its isotonic region, within which the point might be positioned without affecting stress. Obviously, the more data there are, the smaller such isotonic regions become. It remains the case, however, that whereas the overall (global) structure of a configuration is stable and reliable, local information (information about close or adjacent points) is not stable because of this freedom to move each point within its isotonic region.

We shall see that it is rarely possible to interpret the whole of a configuration. Rather, we will be discovering structure within parts of it. We want to be careful not to place too much stress (pun intended) on these local instabilities. Sometimes we can detect which points are most unstable in their location—either from studying the point contribution to stress (see 3.6.2.1), and/or by making several runs from different initial configurations and comparing the final configurations to see which points tend to change location. But in any event, analysis which relies upon small differences of location is not recommended, since it will almost certainly capitalise upon non-unique and unstable characteristics of the solution.

4.2 Internal and External Interpretation

Arrangements of points in a space do not normally exhibit any self-evident structure; we have to bring additional information to the task of interpretation. Two aspects will concern us particularly: pattern and meaning. Patterning refers to the way in which points are located and related quite independently of what they may 'mean'. This may be evident by inspection or discovered by analysis. It is not usually difficult, for instance, to identify sets of points forming a straight line, or a circle, or a parabola, or even a set of discrete clusters—but it is more difficult to pick out general directions, or overlapping clumps, or even a 2-dimensional plane in a 3-dimensional space.

The 'meaning' of a configuration is a more complex matter. Once labels are attached to the points we bring all sorts of other information into play—what we know about the objects, what connotations they have for us, what subjects said about them and so on. As these meanings are put together we begin to recognise more subtle relationships. In fortunate circumstances, hitherto unsuspected characteristics of the data may then become apparent.

There is no procedure that will automatically detect structure in a configuration. The procedures described below will only assist the user to set about the task of identification in a fairly systematic manner, but there is no guarantee that the types of structure identified will be the most significant ones or relevant in any particular analysis.

It is worthwhile at this point to distinguish between *internal* and *external* methods of interpretation. In internal analysis only the original data are used in interpretation, whereas additional information is employed in the external case.

If we are to use only the original data in our interpretation we have two broad alternatives. Aspects of the original data may be represented in a graph-theoretic way, as line-segments within the configuration. This is a useful method by which simple structures may be identified. Alternatively, we may submit the same data to a clustering analysis and use this to interpret the scaling solution.

On the other hand, interpretation is often made easier by using information about the stimuli obtained independently of (externally to) the scaling itself. For instance:

in scaling subjects' judgments of similarity between nations one might use economic and political information on the nations concerned (cf. Wish 1972);

in scaling judgments of psychological stimuli (subjective loudness, brightness, colour saturation and hue), one would use information on the physical or objective variables involved (Carroll and Wish 1974);

in scaling judgments of personality-trait words, one might use known semantic properties and/or subjects' own ratings of the general properties of the words (Rosenberg and Sedlak 1972).

A closely related source of information for interpreting the meaning of a configuration, at least when the original data are obtained from human subjects, is what they themselves say when generating the original data. This is a much under-used resource, since a fair proportion of studies encourage subjects to verbalise as they produce data. Subjects' comments can be inspected and analysed in terms of their general semantic or particular substantive content and then related to features of the scaling representation.

4.3 Internal Methods of Interpretation

In this section we consider three aspects of configurations which have received particular attention: the *dimensions* (orthogonal axes which span the multidimensional space), the *regions* (or concentrations or high density of points, differentiated from others by empty regions) and the *simple structures* (identifiable one- and two-dimensional simple patterns).

4.3.1 Spanning dimensions

A set of orthogonal reference axes is necessary to locate any set of points. As we have seen, in the case of simple Euclidean distance the orientation of the axes is arbitrary yet a good deal of effort has been devoted, especially in the factor analysis tradition, to 'identifying' or naming them.

The important characteristics of a dimension, within this tradition, are that it represents a higher order organising construct ('factor'), which can be thought of as varying continuously and is bipolar (i.e. varies in both a positive and negative direction), and defines a major pattern of variation in the data (cf. Rummel 1970, ch. 21).

Having decided upon a set of reference axes, the dimensional analyst first separates out the objects or stimuli with the most extreme (positive and negative) co-ordinates on each dimension compared to those nearest zero, in order to establish which are the relevant and which the irrelevant objects for identifying the dimension concerned. Then the objects with the highest co-ordinates are compared to those with the lowest co-ordinates in order to identify the bipolarity of the dimension, or the contrast involved, if such there be.

The process of naming the dimension is by its very nature difficult to systematise. In effect, the researcher is performing a cognitive task analogous to that which social scientists often ask of their subjects (see 2.1). That is, 'In what way(s) do the high and low points differ?' Also, 'What property/properties do these points share, which others (in the zero position) do not possess?' The answer—the term used to label the dimension—depends in part on the researcher's verbal or conceptual abilities or on the accessibility of Roget's *Thesaurus*. In factor analysis and MDS, the most frequently encountered set of reference axes are principal components (see 3.5.5).

While it is useful to find out what the direction of maximum variation in the configuration actually is (the first principal component), there is no reason to suppose that it will be substantively significant or meaningful. On the other hand it may direct the user's attention to a readily identifiable, significant, general variable or factor underlying the configuration—general intelligence, general occupational prestige, the overall size of specimens, the left-right political continuum, the evaluative factor in word connotations, tough-tender-mindedness in personality studies—all have been much-heralded primary dimensions of variation.

What of the other dimensions of variation? It depends principally upon whether the user wishes to keep the dimensions orthogonal to each other. If so, a number of techniques exists for finding a set of reference axes which will often lead to a more interpretable set of dimensions, by (rigidly) rotating the principal components to a new position where, for instance, the co-ordinates of the points on a given axis tend to either unity or zero* (a 35° counter-clockwise rotation of axes in Figure 1.1

*Kaiser's 'varimax' rotation criterion. See Rummel 1970, pp. 391–3 or Maxwell 1977, p. 54 et seq. for a simple introduction and exposition.

above comes close to satisfying this requirement). If the MDS analyst wishes to pursue the identification of 'best' axes, then the detailed technology developed in the factor analytic tradition (cf. Rummel 1970, chs. 14–21) may be used. On any account, the naming of dimensions on purely internal criteria can be a hazardous and subjective undertaking, at least for the basic MDS model.

4.3.2 Graphical interpretation

A configuration may be interpretable in differing ways in different regions and there is no guarantee whatever that one particular type of structure will best describe the entire configuration. But if we hope to detect different types of local structure we shall need a procedure that is sufficiently general to cover other methods as special cases. A technique that has been found to be very serviceable in detecting different types of local structure is graphic analysis (Kendall 1971a; Waern 1971).

The basic idea is to represent the highest similarity values in the data by drawing a line in the MDS configuration between each pair of objects involved. This can be done either by deciding upon a cut-off value (say the top quartile of data, or all similarity values greater than 0.70) or by ordering the data from the highest to the lowest similarity values. If the data are ordered by size, the link between the highest similarity pair is drawn in the configuration first, then the next highest pair is linked and so on, until the researcher has sufficient information about the local structure. In this way the growth of various types of structure becomes apparent as the researcher moves down the ordered data list. These structures may include:

(i) *chains:* successive links which form a path through the configuration. The path may be (approximately) linear, signalling a vector or dimension-like property, or non-linear, a parabola, circle or some other simple regular pattern, or even an irregular, zigzagging, but connected, sequence.

(ii) *clusters:* links which occur within particular regions and build up to form locally-connected subgraphs or clusters based upon a few points.

In many instances both types of structure may appear, producing a linked set of clusters and often a residual set of isolated points.

4.3.2.1 Sequences and seriation

The historical scientist is naturally interested in inferring or detecting the time-sequence of a set of objects or events. This process is usually referred to as 'seriation' or 'ordination' (see Renfrew 1976, ch. 2; Hodson 1971, pp. 173–290; Hubert 1974). A good example of the success of the graphical method in finding a sequence is reported in Kendall's paper (1971b) in which he analysed an 'abundance matrix' consisting of co-occurrence counts of various artefacts such as type of pottery and jewellery common to pairs of graves in the neolithic cemetery of the la Tène culture at Müssingen-Rain. By using graphical analysis as described above, he showed that a chaining emerged whose shape led to its being described as a 'horseshoe'. Kendall hypothesised that the sequence of the points along the horseshoe indicated the historical order of the burials. This hypothesis was subsequently confirmed in an independent study by Hodson.

Why does a continuum (such as, here, time) become distorted in this way? There is no definite answer, but the evidence suggests that it is due in part to the fact that some permissible monotone transformations of the straight line generate a semi-circular MDS two-dimensional configuration (Shepard 1974, van de Geer, 1971, pp. 239–42). There is also increasing evidence (Kendall 1971a, p. 225) that it may be due to data collection procedures imposing an upper ceiling on values which the dissimilarities can take.

Thus, for example, in a rating exercise where only five or seven categories are allowed, the 'totally dissimilar' or 'totally unalike' category restricts the subject's ability to distinguish between very dissimilar and the very, very dissimilar. This, in turn, restricts the number of distinct values in the data matrix and the ability of the non-metric procedure to make the distance between points at opposite ends of the continuum as large as they 'should' be, thus producing the noticed horseshoe shape. This phenomenon can be overcome, i.e. the horseshoe sequence straightened out by treating the data as being only *locally* Euclidean, or by assuming that the largest distances are of least (or no) value in determining a solution, and relying only on proximate information. This may be done by using continuity mapping (5.2.2) or local monotonicity (5.2.1.1). The apparently simpler expedient of scaling in one dimension to recover such structures is usually misleading and is not recommended. This is partly due to the fact that highly irregular non-linear continua occur. As Shepard (1974, p. 386 et seq.) comments:

> In analyses of many different sets of data that were known to be basically one-dimensional, I have found that two-dimensional solutions, when attempted, characteristically can assume either the simple C-shape or the inflected S-shape ... Evidently, by bending away from a one-dimensional straight line, the configuration is able to take advantage of the extra degrees of freedom provided by additional dimensions to achieve a better fit to the random fluctuations in similarity data.

Users should therefore be on their guard against using one-dimensional solutions to recover an ordering and should use graphic procedures to help detect strong non-linearities of this sort.

Perhaps the best known, and independently replicated, instance of a C-shaped continuum in a 2-space is of adjudged musical intervals: the intervals separating the points correspond very well to the number of intervening semitones when they are projected onto the horseshoe, but no linear dimensional interpretation is possible (Levelt et al. 1966; Shepard 1974, pp. 386–7).

4.3.2.2 Simple structures

Other simple graphical structures have been studied and identified in scaling solutions, either from empirical patterns in the data or, in the Guttman tradition of facet theory (see 4.5) from the structure of the original mapping sentence (Lingoes and Borg 1979; Maimon et al. 1980). Among the more commonly encountered structures, in Guttman's terminology, are:

the simplex (a chain, a linear or non-linear dimension or sequence);

the circumplex (a circular arrangement of points);

the radex (a combination of the simplex and the circumplex), consisting of two or more concentric circles with lines emanating from the centre, dividing the circles into sectors and thus the sheep from the goats. In three dimensions a set of stacked radexes is termed a *cylindrex*.

A simple example of facet-theoretic interpretation occurs in Levy and Guttman (1975) and a simplified version of the relevant mapping sentence is presented in section 4.5. The two-dimensional MDS solution for fifteen of the items on a USA sample of subjects is presented in Figure 4.1 and includes its radex interpretation. In such an interpretation the configuration is divided up into a number of regions within which particular object points possess a particular characteristic, or facet. In this instance, where the radex defines a number of sectors, item 15, life in general (LG), functions as the centre of the circle, the facet state *vs* resources form the inner and outer circular bands and the facet 'area of life' divides the circle into eight sectors corresponding to those of the original mapping sentence.

A similar simple structure was found in a re-analysis (Coxon 1974) of the Bollen-Delbeke data (Delbeke 1968) on preferences for families consisting of different size and composition (of boys and girls). Using the basic (distance) MDS model, the two constituent dimensions (number of boys and number of girls) were identifiable as two somewhat distorted lines, reflecting the empirical fact that people tend to prefer mixed family compositions. Using the vector model (see 5.3.2 below) for analysis of the same data, a radex structure was apparent. In this case, the inner and outer circular bands represented mixed *vs* unmixed composition, and the sector lines divided points in terms of the overall size of the family. These data and their analyses are discussed at greater length later, in sections 6.2.2 and 6.2.3.

A fascinating example of a cylindrex pattern ('stacked' circles) occurs in Heider

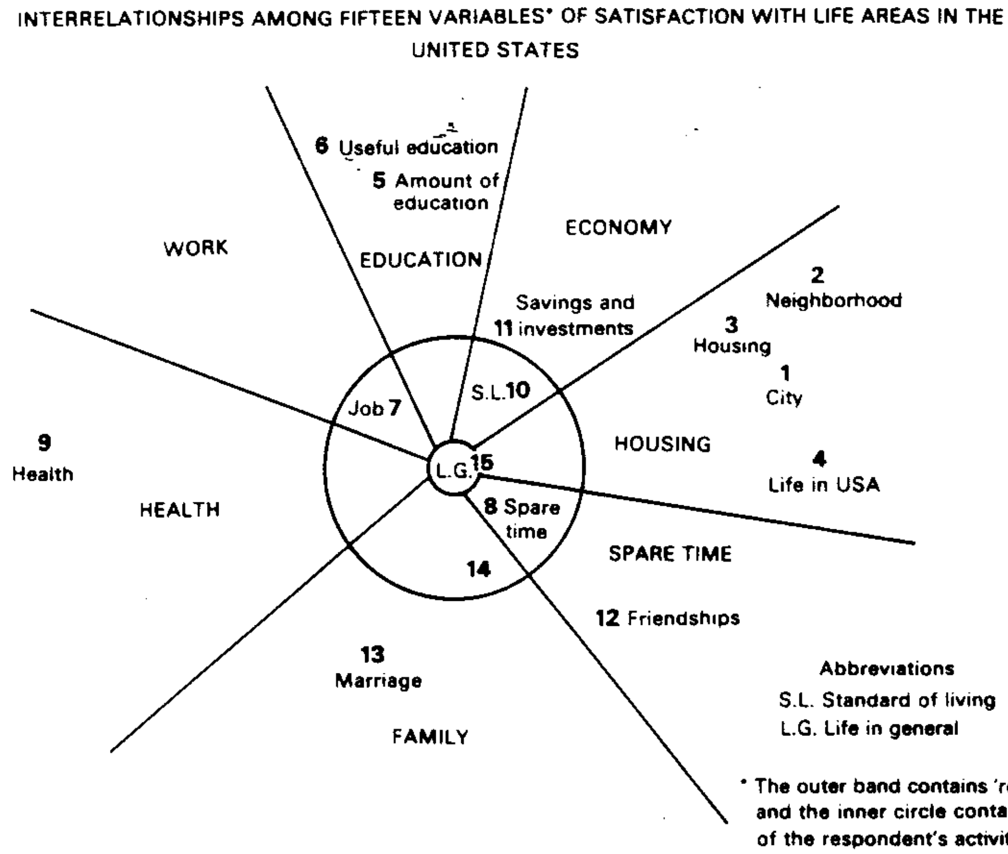


Figure 4.1 *Interrelationships among fifteen variables of satisfaction-with-life areas in the USA*

and Olivier's analysis (1972) of the structure of colour naming and memory in two markedly different languages. (The so-called Whorf hypothesis asserts that categories of naming used in different languages in some sense 'cause' different perceptions of reality, hence the interest of this study.) In this case, as in other studies of colour perception, the two basic dimensions of the space (brightness and hue) have in the MDS configuration become 'wrapped round' the brightness axis to form a circular pattern in 3-space, bringing together the red and purple ends of the hue dimension. The substantive interest in the Heider and Olivier study is to see whether the restricted categorisation of the Dani (a New Guinea people who use only two basic colour terms) affects their ability to recognise and identify different colour chips, compared to an English-speaking sample. Although the effects of restricted categorisation can be seen in the naming task configurations, it seems that retention of colour image in memory is unaffected by the very considerable cultural and semantic differences of the two languages. In both cases, colours of differing hue but the same brightness form a circular pattern, and the circles stack at different levels of brightness.

4.3.3 Regions of high and low density of points

Casual inspection and/or graphical analysis of an MDS configuration usually reveals that the points are not evenly distributed over the space. Rather, points tend to clump or cluster together, reflecting their high similarity, and are separated from other clusters by empty or sparsely-populated regions. That is, there exist subsets of points the relationships within which are stronger than those between them (Lingoes 1977, p. 116). Users may wish to check whether these clusters display any more formal structural properties.

A whole family of models, as extensive as those of MDS, exists for identifying and relating clusters of similar objects, namely cluster analysis (Wishart 1978; Everitt 1978). However, cluster analysis gives no information on the *extent* of separation of the clusters, and for this reason it is often advantageous to combine clustering analysis with MDS, which, of course, represents distances directly.* The best strategy is therefore to analyse the data *separately* by both a clustering and an MDS model and then represent the clusters *within* the MDS solution configuration. This is done in a 2-space by drawing a closed contour around the points contained in a cluster. The regions so enclosed will represent areas of high density, and the extent of their dissociation will be the distance in the configuration. This is done in a two-dimensional MDS plot: extension to three dimensions is usually not impossible, but poses problems in graphical portrayal.

Two varieties of clustering are extensively used in conjunction with MDS analysis—*hierarchical clustering schemes* (Johnson 1967, implemented as HICLUS in the MDS(X) series) and overlapping or *additive clustering* (Shepard and Arabie 1979, implemented as MAPCLUS; see section 8.2).

*Some data analysts recommend scaling followed by a clustering of the resulting *distances*. This practice is not be recommended since it capitalises on the weaknesses of both methods. As we have seen (4.1), MDS solutions are least stable in their fine-grain local structure, on account of the existence of isotonic regions, but this is the very information from which clustering initially proceeds and which significantly determines the final clustering.

4.3.3.1 Hierarchical clustering schemes (HCS)

An hierarchical clustering scheme takes a matrix of dissimilarity measures between a set of objects and represents the objects as being gathered into clusters on the basis of this information. It describes not one clustering but rather (for p points) p different clusterings, referred to as *levels* of a single total hierarchical *scheme*. At the highest level, all the objects are contained in one cluster, at the next highest there are two and so on until, at the lowest level, there are as many clusters as there are points. The defining characteristic of a hierarchical scheme is that once a point is incorporated into a cluster at a lower level it may not 'leave' that cluster at a higher one. Thus the clusters form a hierarchical scheme in the sense that each level is a special case of the next highest. We now consider in some detail the method of hierarchical clustering.

Data (similarities)

	A	B	C	D	E
A	0	4	2	2	4
B	4	0	4	4	3
C	2	4	0	(1)	4
D	2	4	(1)	0	4
E	4	3	4	4	0

(Stage 0)

	A	B	C,D	E
A	0	4	(2)	4
B	4	0	4	3
C,D	(2)	4	0	4
E	4	3	4	0

(Stage 1)

	A C,D	B	E
A,C,D	0	4	4
B	4	0	(3)
E	4	(3)	0

(Stage 2)

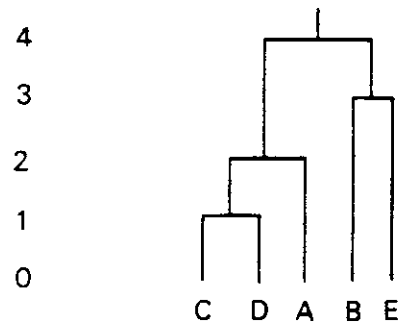
	A C,D	B E
A,C,D	0	(4)
B,E	(4)	0

(Stage 3)

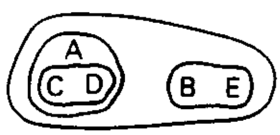
```
Level          Objects

          C   D   A   B   E
 0        .   .   .   .   .
 1        XXXX    .   .   .
          XXXX    .   .   .
 2        XXXXXXX     .   .
          XXXXXXX     .   .
 3        XXXXXXX    XXXX
          XXXXXXX    XXXX
 4        XXXXXXXXXXXXX
          XXXXXXXXXXXXX
```

(i) HCS representation



(ii) Dendogram representation



(iii) Contour representation

Figure 4.2 *Illustrative example of the HCS procedure and forms of representation*

The Method

Stage 0 The process of clustering begins by inspecting the original data matrix of dissimilarities and identifying first the most similar (or least dissimilar) pair of objects (C and D) and then merging them into a cluster. We now have the closest cluster of two points: (C, D).

Stage 1 Points C and D are from now on treated as a single *object* and the data matrix is reduced by removing the row and column of C and D and substituting one representing the cluster (C, D). In this example, the dissimilarity between C and each other object is the same as that between D and the same object—e.g. $\delta(C, A) = \delta(D, A) = 2$. (Normally this will not be the case.) The smallest entry in the reduced matrix, the currently most similar link, is now identified, and turns out to be between the cluster (C, D) and A. So the new structure consists of the dense cluster (A, C, D) and a set of unlinked points.

Stage 2 The new reduced matrix consists of the cluster (A, C, D) and points B and E. The smallest entry is now between B and E. This pair now form a new distinct cluster; the structure at this stage consists simply of the two clusters: (A, C, D) *vs* (B, E). A new reduced matrix is formed.

Stage 3 In this final stage, the remaining two entities—the two clusters (A, C, D) and (B, E)—are merged, forming the final clustering of all the points.

The process of hierarchical clustering—forming clusters at decreasing levels of compactness—gives considerable insight into the regions of the space. In this case, we can see that the basic contrast is between the (C, D, A) and (B, E) cores of clustering.

As in other areas of data analysis, especially block modelling of social networks (see White et al. 1976 and Breiger et al. 1975) 'holes', the empty areas, frequently turn out to be quite as significant as the clusters, and both aspects have to be represented in any structural analysis. Empty regions represent two types of significant information: differentiation or dissociation between clusters on the one hand, and/or the significant absence of objects on the other hand, which might mean that certain stimuli have been neglected or overlooked in a study or that no objects actually exist which have a particular combination of attributes.

Clearly, the most significant *clustering* information is contained in the initial stages, and the most significant *dissociation* information is contained in the later stages of a clustering.

In the above example there was no ambiguity in defining the distance between a newly-formed cluster and existing objects (clusters or points), but this will not usually be the case. Consider the simplest case where we have a cluster formed of two points A and B and a third point C. There will be two distances, namely those between A and C, $\delta(A, C)$, and between B and C, $\delta(B, C)$; and we have to decide how we are going to use these to define the distance between (A, B) and C, that is $\delta((A, B), C)$. If we want the procedure to produce identical clustering schemes when the data are monotonically transformed we cannot take the obvious step of averaging $\delta(A, C)$ and $\delta(B, C)$. Johnson (1967) suggests two contrasting ways of defining this distance in this instance:

The *maximum* method (otherwise known as the diameter or complete link method) defines the distance $\delta((A, B), C)$ to be the *maximum* of $\delta(A, C)$ and $\delta(B, C)$.
The alternative *minimum* method (also known as the connectedness or single-link method) conversely defines the distance between the new cluster and the extraneous point to be the *minimum* of the distances between the extraneous point and each of the points in the cluster.

When the data satisfy the ultra-metric inequality (see 6.1.6) and are therefore perfectly representable as an HCS, the two methods produce identical hierarchical clusterings. Otherwise, the two HCSs will differ—often not markedly, but sometimes significantly.

The maximum (diameter) method picks out the largest distance within a cluster as 'the' distance and seeks to minimise the diameter (largest distance between the objects) within a cluster. This tends to produce a fairly small number of compact clusters.

The minimum method, by contrast, selects the smallest distance as 'the' distance and seeks to minimise the largest link needed to produce a chain or connected path between the objects. It tends to produce rather a large number of broken clusters and is often marked by chaining—the continued addition of a single element to a cluster.

In practice, the minimum method is usually to be preferred to the maximum method in exploring the hierarchical structure of a set of data (although both methods should be inspected to see how far the data may legitimately be represented this way.* The chief use of the HCS procedure is to examine not only relatively dense 'local' structure of highly proximate points (the lower levels of the clustering) but also the open or 'global' structure of spaces which separate or dissociate the clusters (the highest levels).

Hierarchical clustering then, possesses a number of useful characteristics:

it presents not one, but a whole series of linked clusterings of increasing density, from a 'clustering' where each point is a separate cluster to the one where all the points are in a single cluster;

it includes two commonly used types of clustering as special cases and therefore gives the user some idea of how well the data fit the assumptions of the clustering model;

the HCS procedure is non-metric, in the sense that any ordinal rescaling of the data will produce identical results.

4.3.3.2 Clustering in high-dimensional space

The simple representation of HCS solutions within MDS configurations is only really feasible in two- or, at most, three-dimensional space. What if the user has a higher-dimensional solution and wants to gain some insight into the differential density of points in that space? An ingenious procedure is suggested by Andrews (1972; also see Everitt 1978, pp. 81–6) to represent each point as a wave form.

*Holman (1972) has shown that a set of errorless data will never perfectly satisfy both the Euclidean distance model and the hierarchical model, but will always satisfy one of the models to some extent.

Given a set of points in r-dimensional space, each point x is defined by its r co-ordinates: $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_r)$. It is then represented as a Fourier series function of the form: $f_x(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \ldots$, and the function is plotted for $-\pi < t < \pi$. The proximity of points can then be studied in terms of the similarity of the wave forms:

> If some plotted functions form a band by remaining close together for all values of t then the corresponding points are close together in the Euclidean metric.
> (Andrews 1972, p. 133)

Because the wave function preserves Euclidean distances the points which are close together have wave forms that have highly similar wave shape, whereas distant points have wave functions whose shape is different.

Because the wave function is affected by *all* the dimensions, the salient features of a high-dimensional configuration can be studied, and the procedure is therefore very useful for detecting isolates or outliers (which have markedly different wave forms) and clusterings (which have markedly similar wave forms).

4.4 External Methods of Interpretation

So far we have made use only of the original data in seeking to detect structure in the MDS solution.

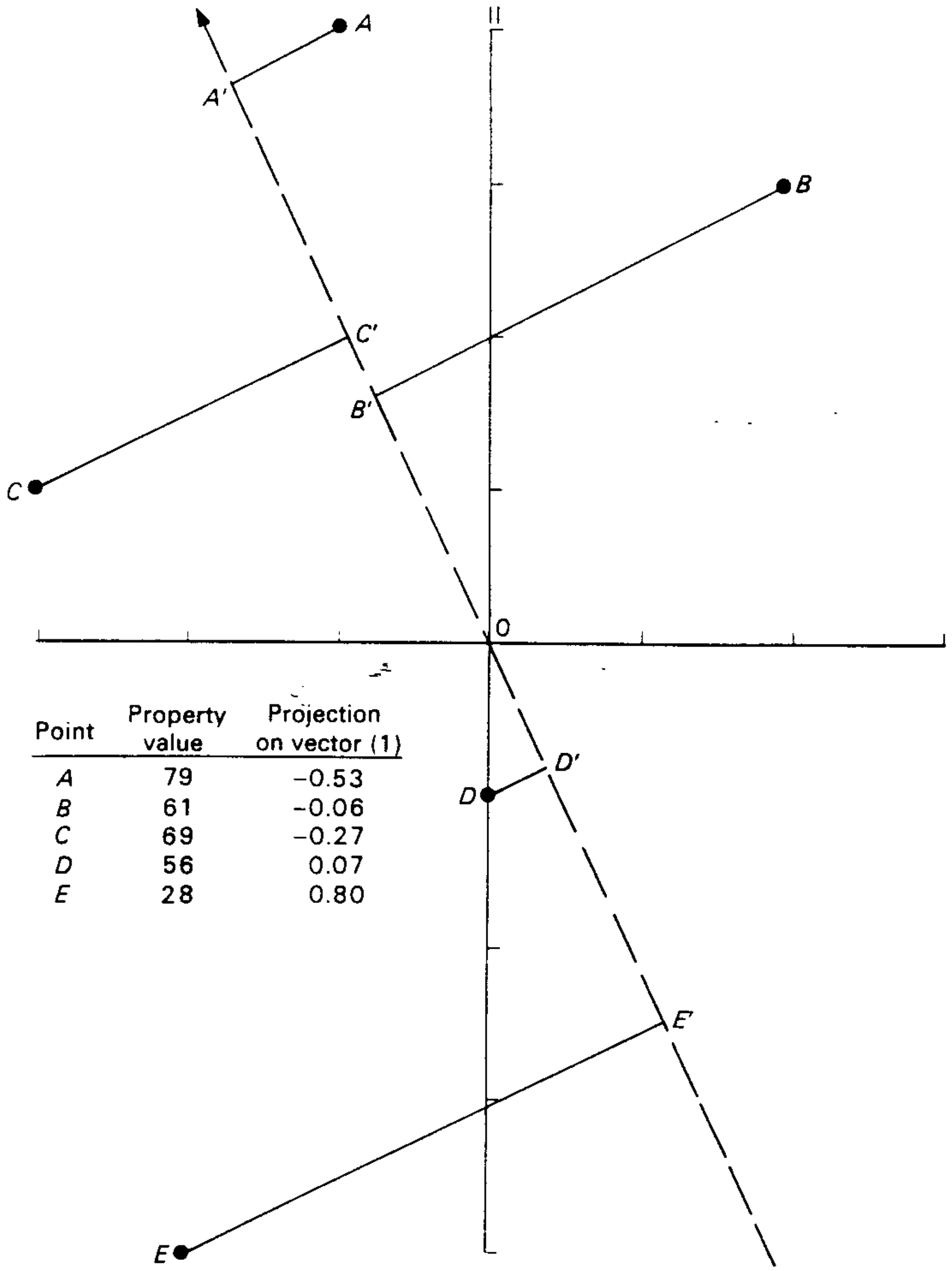
When the researcher possesses additional external information about the points in the configuration, the task of interpretation is made much easier. These external variables, or 'properties' as they are often called, may come from a variety of sources. They may be relevant physical characteristics; they may be judgments made by respondents separately from the judgments used in the scaling; or they may simply represent the hunches or hypotheses of the researcher about the nature of the configuration. In any event, each property is assumed to consist of a set of numerical (interval or ordinal) values of the variable concerned, one for each point in the configuration.

A number of procedures exists for representing or 'embedding' each property within the already-obtained configuration, in a simple and easily recognisable manner. Two commonly used forms are as a *vector* and as a *point*. In essence a vector is a line drawn through the solution space and pointing the direction in which higher values of the property occur. Thus if our points were geographical sites, one relevant property might be 'northness', i.e. each point would have a value which was its north latitude. The property would then be fitted into the configuration map so that it was directed towards the pole. Such a representation would be adequate for points within areas of the northern hemisphere such as Scotland or the USA, but if the configuration actually contained the north pole, such as one consisting of sites in North America and Asia, in fact a map drawn from a vantage point above the pole, then the property of 'northness' would have to be represented not as a vector but as a *point* at the pole from which this property 'north' would decrease uniformly in all directions. Notice that representation in terms of vector or point depends on the characteristics not only of the property but also of the configuration.

These two forms of representing external information in a configuration are illustrated in Figure 4.3, with reference to the same 5-point configuration. Figure

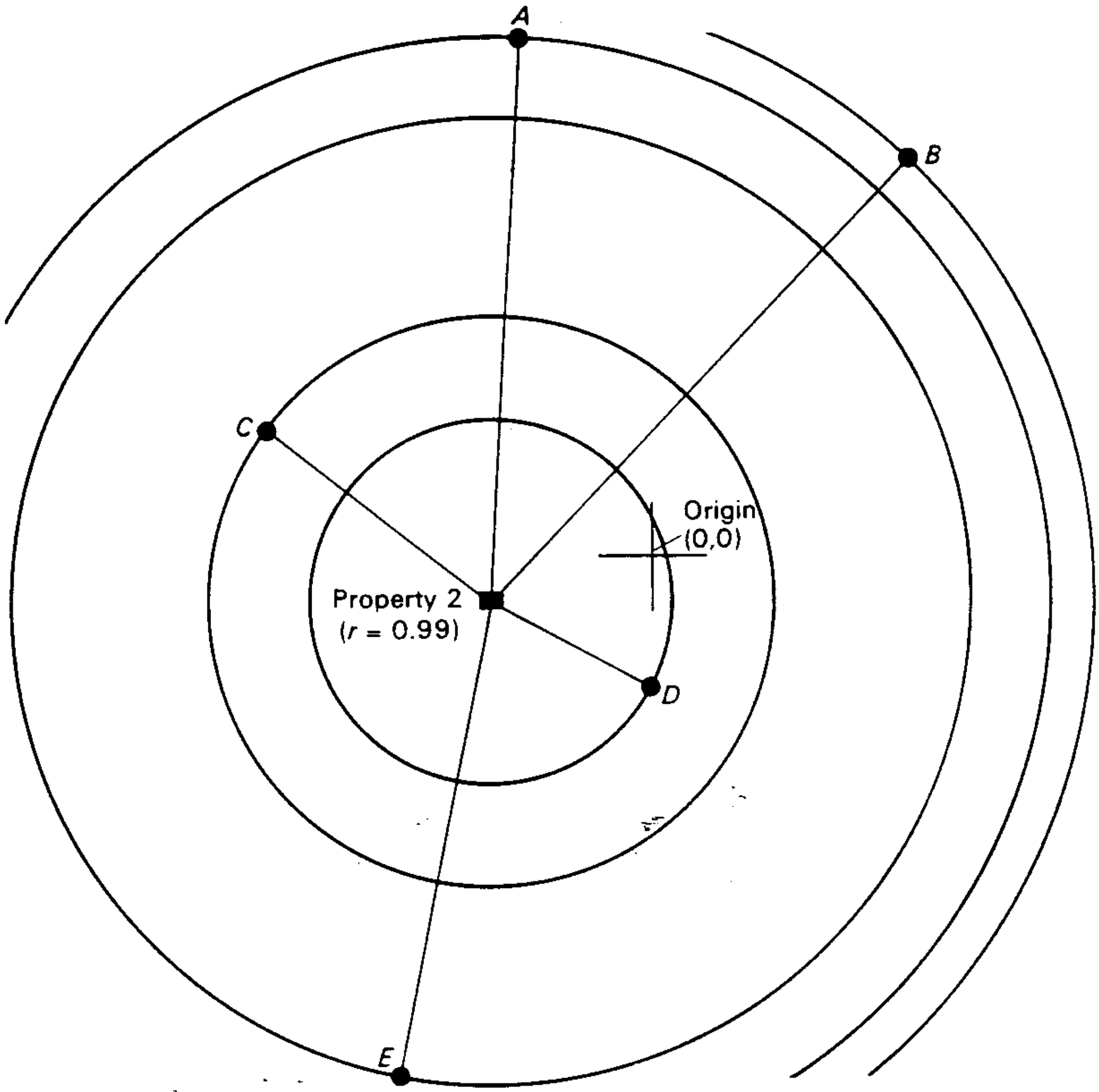
4.3a represents a property of the points as a vector pointing in a north-easterly direction as the property values increase. Figure 4.3b represents another property in the same configuration of points, with the highest occurrences of the property somewhat left and below the origin of the space and systematically declining in all directions from this 'peak'.

These two ways of representing an external property within a configuration are now taken up in turn.



Point	Property value	Projection on vector (1)
A	79	−0.53
B	61	−0.06
C	69	−0.27
D	56	0.07
E	28	0.80

Figure 4.3a *Representation of an external property as a vector*



Point	Property value	Distance from point (2)
A	52	1.09
B	63	1.26
C	28	0.28
D	18	0.12
E	45	0.83

Figure 4.3b *Representation of an external property as distances from a maximal point*

4.4.1 Properties as vectors

Locating a single property in a space as a vector is similar to finding a dimension in the space, except that a vector is unipolar and not bipolar as is a dimension. A property consists of a value for each point in a configuration and the aim of the procedure is to position the vector in the space (like an axis) so that the projections onto the vector (like co-ordinates on the axis) correlate with the property values in some well-defined sense.* Such a procedure has several uses: it accurately locates a

*Clearly, the match between the original property values and the projections on the property vector should be as close as possible. Current options in MDS(X) include maximising *ordinal* fit (PREFMAP IV with monotone option), *linear* fit (PREFMAP IV with linear option, or PROFIT with linear regression) and *non-linear 'continuity'* fit (PROFIT with continuity option). See section 6.2.1.

direction in the space for each property; it gives the researcher some assurance as to how accurate are her hunches about properties of the space, and it provides a useful way of mapping additional information into the space.

When more than one property is fitted into the same space, attention is focussed additionally upon the relationship between the fitted vectors. The basic information is quite simply the angle separating them. Thus a right angle represents independence (zero correlation), in which case they are equivalent to a set of a axes; 180° represents perfect *negative* association; and a 45° angle represents a linear correlation of 0.707 (cos 45) and so forth.

Obviously, property-fitting can also be used to *identify* axes of a configuration by inspecting how close an external property comes to pointing in the same direction as the axis concerned. An example is presented in Carroll and Chang (1969, pp. 290–3), where they identify the three dimensions of a configuration obtained from scaling judgments of a set of tones as modulation frequency, modulation percentage and modulation waveform, to within 8°, 5° and 14° of the axes of the configuration.

An example

An instructive example of the uses to which vector property-fitting can be put is Rosenberg's studies of implicit personality theory (Rosenberg and Sedlak 1972a, 1972b). Each subject was given a set of 60 trait names (such as reserved, good-natured, submissive, humourless, etc.) and asked to sort them into groups, each of which was to represent a different person that they knew. A co-occurrence measure was used as a basis for constructing a dissimilarity measure between pairs of traits, and then scaled, yielding a two-dimensional solution, with $stress_1$ of 0.09. The subjects had also been asked to rate each of the 60 traits in terms of 5 general semantic differential scales (Osgood et al. 1965): hard/soft; good-intellectual/bad-intellectual; active/passive; good/bad; good-social/bad-social. The averaged ratings formed five properties, which were then fitted into the configuration as vectors (see Figure 4.4). Rosenberg and Sedlak's interpretation well exemplifies the use of property fitting and merits extended quotation:

> Five properties are obviously not needed to interpret a two-dimensional space. Moreover, there are alternative pairs of properties, all with high R (linear correlation) values, which can be used to interpret this space.
>
> It is possible, for example, to interpret the two-dimensional space in [the] figure [4.4] with the two general semantic differential factors, good-bad and hard-soft. Also, if we consider the three-dimensional solution, the R value for active-passive increases to 0.585 ($p < 0.001$), and the angle between the fitted axes for the three semantic differential factors are:

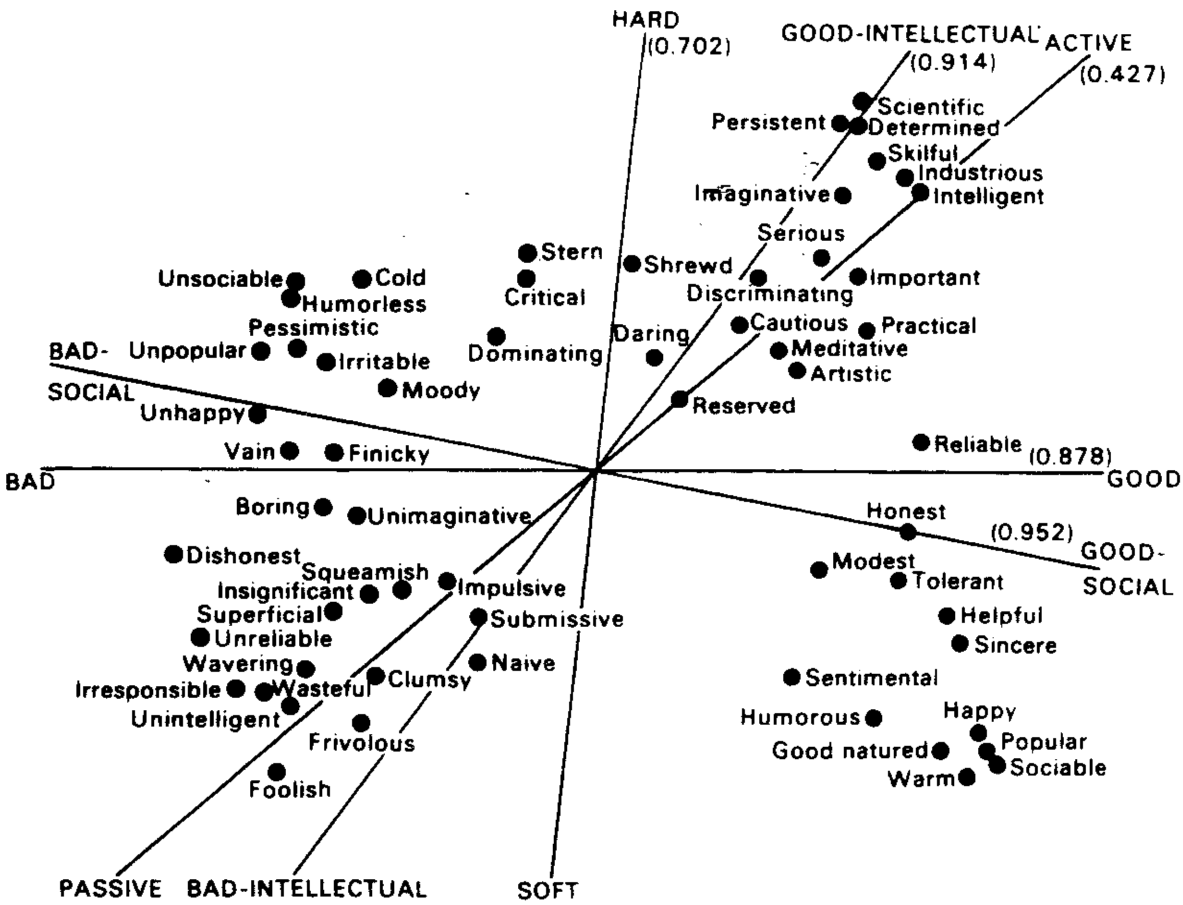
	good-bad	hard-soft
hard-soft	83°	
active-passive	92°	76°

> Thus, the results from the three-dimensional solution support the presence of the three semantic differential factors in personality perception with each factor more-or-less orthogonal angles (≅ 90°) to the others.
>
> An alternative interpretation, at least for the two-dimensional space in [the] figure [4.4] is the use of the two descriptive-evaluative properties, social good-bad and intellectual good-bad. It is interesting to relate this interpretation to Hays' (1958) findings that the extreme traits on one of his rank-order

dimensions were warm and cold (social good-bad?), and on the second rank-order dimension they were intelligent and stupid (intellectual good-bad?).

These two descriptive-evaluative properties are not orthogonal, however, either in the two-dimensional (65°) or the three-dimensional (67°) spaces. Nevertheless, while orthogonality is convenient, it is not a necessary feature of an acceptable interpretation of a space. Nor are the two sets of interpretations of the two-dimensional solution incompatible. The heuristic distinction between connotative and denotative meaning is relevant here. The interpretation based on hard-soft and good-bad might be thought of as concerned with connotative meaning, whereas the interpretation of the space in terms of social and intellectual desirability is concerned with denotative meaning. A multi-dimensional analysis of the basic denotative meanings of traits may not result in orthogonal co-ordinates. Indeed, intellectual traits are likely to be perceived by college students as relevant for social activities.

(Rosenberg and Sedlak 1972b, p. 253)



2 dimensional configuration of the 60 traits scaled by Rosenberg *et al.* showing the best-fitting axis for five properties. Each number in parentheses indicates the multiple correlation between projections on the best-fitting axis and property values. Reproduced by permission.

Figure 4.4 *Rosenburg's trait configuration interpreted by 5 property vectors*

When a property is represented as a vector the assumption is that the property is increasing across the space in the direction indicated by the vector and that the limit of the property is at infinity. Thus the lines joining points equidistant from the limit or ideal point are parallel to the vector. If the limit is brought in from infinity to the boundary of the space, i.e. the property is increasing but with a finite bound, then the equidistant lines assume the familiar convex shape of the indifference curves—iso-preference contours—of micro-economic analysis. If the limit or ideal point is brought within the boundary of the space, i.e. becomes not only finite but accessible, then the iso-preference or iso-similarity lines become circles around the point. As in our example of the property 'north' above, such a representation may make assumptions not only about the characteristics of the property but rather about the characteristics of the space. Thus, whereas in the case of vector representation the vector is positioned so that the perpendicular projections (co-ordinates) onto the vector of the points matches the property values, in the case of the point representations it is the distances from the 'property point' to each of the points in the configuration which are matched to the property values.

4.4.2 Properties as points

If the user collects data which relate a new object to each of the existing ones, and then interprets the information as distances, it is straightforward to locate it as a point in the original configuration (the SSAM and PREFMAP III programs allow just such an option). A useful application occurs when the user wants to position some new points in a configuration that is either already known or where the information for some points is more reliable than for others.

An example which illustrates such use is Tobler and Wineberg's study (1971), based upon the co-occurrence of the names of a number of Bronze Age merchant colonies in Cappadocia on a set of some 800 cuneiform tablets. The co-occurrence frequencies were taken to be a function of the size of the colonies and of their geographical separation, and the data were scaled in two dimensions. Unfortunately—but hardly surprisingly—the location of the great majority of the 65 Bronze Age towns finally used in the analysis were not known. But had the location of a significant fraction been known, they could have been treated as a known, fixed, geographical framework. The co-occurrence information for the remaining towns could then have been treated as a set of external properties and located as points within the known configuration. In this way it would be possible, in principle, to identify the location of colonies which had subsequently disappeared. (This is also illustrated in Kendall's famous paper (1971b), 'Maps from marriages').

Probably the most common use of property-fitting as points occurs in preference studies. In this case, a subject's numerical evaluations of a set of objects are located as a point of maximum preference in a configuration which has been previously obtained from judgments of similarity of the same objects. (This use is discussed in 5.3.3.1).

In all these instances—adding new objects to a configuration, mapping subjects' preferences as 'ideal points' or representing an external property as a point as an aid to interpreting a configuration—the basic principle is the same. The additional information is viewed as giving a set of relative distances from the new point to all the existing ones. The task of the scaling program is then to position the point so that it best reproduces those distances (or their rank order). To date, little use has been made of this way of representing a property as a 'high point' in the configuration, but it has considerable advantages.

4.5 Facet Theory as a Framework of Interpretation

There exists a methodology which views the whole of the measurement process, up to and including the interpretation of scaling solutions, as an integral whole. This is Guttman's 'facet theory', Space does not permit a detailed exposition here, and the interested reader is referred to Borg (1977). The core of facet theory is the 'mapping sentence' which relates various aspects ('facets') of the population of subjects, or objects/stimuli, and of alternative questions, and maps them into the actual response made. A simplified example relating to a quality of life study might be as follows (cf. Borg 1977, p. 75). (The bracketed and labelled components are referred to as facets, and within each facet the alternatives are separated by a slash (/) symbol.)

> The (*cognitive/affective*)[A] assessment of the (*state of/government's treatment for*)[B] the well-being of the respondent's social reference group (*self/government*)[C] with respect to its (*primary internal/primary social/primary resource/secondary state*)[D] environment concerning a general aspect of life area: (*family/on the whole/economic/...*)[E] according to the respondent's normative criterion for that life area
> IS MAPPED INTO THE RESPONSE-SET: (*very satisfactory/.../very unsatisfactory*)

An actual questionnaire item can be thought of as being an instance of a particular combination of these five facets (A, B, C, D, E). Consider for instance, the following three questionnaire items:

I *'Generally speaking, are you happy these days?'* can be viewed as combining the facets (a_2, b_1, c_1, d_1, e_2).* This specification in brackets is often referred to as a 'structuple'.

II *'In general, how do you evaluate your family life?'* combines (a_2, b_1, c_1, d_1, e_1).

III *'What is your opinion of the way the government handles the economic problems of the country?'* combines (a_1, b_2, c_2, d_4, e_3).

By inspecting the item specifications it is clear that items I and II are inherently most similar, differing only on the E facet (on the whole *vs* economic), whilst item III differs from items I and II and A, B, C, D as well. Clearly, facet theory makes the researcher be more specific about the content of her data collection procedures, and this is itself a good thing. It also makes it possible to inspect the *a priori* or theoretical similarity and relationship of questions *before* obtaining the data. The facet theory approach also formulates a number of useful expectations or

*Subscripts refer to the sequence number of the alternative within a facet.

hypotheses (or 'rules') about the data, e.g. 'The stimuli which are more similar in there facet structure will also be more similar empirically', which would lead us to expect their greater proximity in a distance model solution.

This abbreviated account is sufficient to show that the virtue of facet theory when it comes to interpretation of MDS solutions is that it alerts the researcher to characteristics to be looked for in the configuration, and to the type of structures (e.g. clusters of proximate items) to be expected. To this extent, facet theory can be used to assist the researcher to move beyond simple exploratory investigation towards a confirmatory approach.

In this chapter the simple structures to which Guttman and others have drawn attention have been discussed, but since they were originally developed within facet theory, interested users will profit from inspecting the full context (see Guttman 1971 and Lingoes 1977).

4.6 An Example of Interpretation: Occupational Similarities

An example drawn from one's own experience is most useful in conveying the detail and feel of the process of interpretation. I have therefore chosen work done in the Project on Occupational Cognition with Charles Jones, and concentrated on the interpretation of the basic configuration* of occupational similarities.

Data Pairwise judgments (using a 9-point rating scale) of the similarity of 16 occupational titles were obtained from 287 subjects. Both titles and subjects were selected from a fourfold typology of occupations chosen to contrast level of educational requirements and nature of the job—basically 'People' *vs* 'Data and Machines'.

Configuration Data were scaled in an aggregate (averaged) form, and in unaggregated form. The present example refers to the 2-D projection of the 3-D solution obtained from an Individual Differences Scaling (see Chapter 6) of a subset of 68 subjects' scalings. The basic configuration is that used as the basis for Figure 4.5.

The Resources available for interpreting the configuration were as follows:

(a) *Researchers' 'internal' intuitions and hunches.*

(b) *Data-based summaries for internal analysis*—in this case, the matrix of averaged similarities.†

These are presented in Table 4.1.

(c) *Subjects' verbalisations.* As subjects completed their data ratings, they were encouraged to give the basis of their judgments, and these were either tape-recorded or written on the schedule. In each case, the comments were identifiable as referring to a particular pair of occupations.

(d) *External information from subjects.* Subjects (including a number who had

*The substantive analysis is contained in Coxon and Jones 1978a, chapters 3 and 4 and the technical and methodological material is contained in chapters 2 to 4 of Coxon and Jones 1979. The data on which the analysis is based are available from the SSRC Survey Archive at the University of Essex.

†The original aggregate matrices (mean average, root mean square average and standard deviation of judgements) are presented in T3.8 in Coxon and Jones (1979).

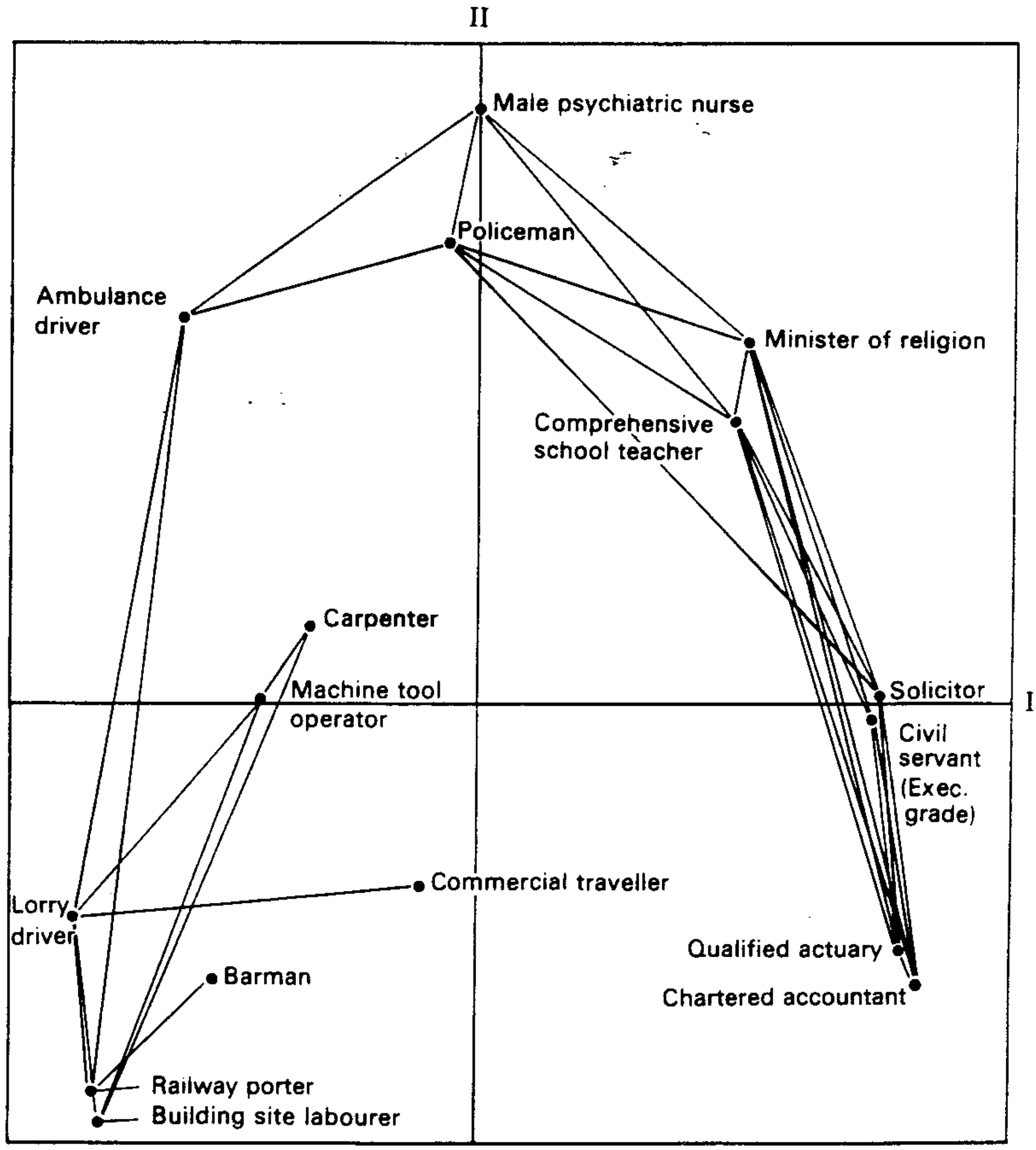
provided similarities data) were also asked to rate and rank the same occupations in terms of a set of characteristics including:

1 Social usefulness.
2 Prestige and rewards they ought to receive
3 Social standing (involving the standard sociological definition in terms of general standing in the community)
4 Monthly earnings (estimated income), together with a measure of
5 Cognitive distance. This last characteristic is not considered further here.

The process of interpretation

(i) *Dimension-naming*

The INDSCAL configuration is by definition already rotated to a non-arbitrary orientation (see 7.2.1), given by the east-west (Dimension I) and north-south



1	(MIN)	–															
2	CST	5.53	–														
3	QA	4.02	4.43	–													
4	CA	4.05	4.47	6.93	–												
5	MPN	4.44	4.53	2.92	2.65	–											
6	AD	3.20	2.76	2.09	2.01	5.13	–										
7	BSL	1.63	1.70	1.46	1.50	2.25	3.01	–									
8	MTO	1.93	2.57	2.11	2.16	2.71	3.25	4.23	–								
9	SOL	5.06	4.63	5.87	6.30	3.14	2.33	1.59	2.00	–							
10	CSE	4.43	4.89	6.10	6.23	3.09	2.35	1.47	2.05	5.71	–						
11	CT	2.91	3.10	2.98	3.28	2.81	3.88	2.50	2.86	3.03	3.06	–					
12	PM	4.27	4.27	2.81	2.90	4.67	5.17	2.47	2.78	4.18	3.69	3.26	–				
13	C	2.28	2.63	2.12	2.25	2.75	3.20	5.03	6.02	2.19	2.13	2.90	3.00	–			
14	LD	1.74	1.83	1.70	1.63	2.33	6.18	5.38	4.25	1.76	1.68	4.87	3.01	3.72	–		
15	RP	1.93	1.85	1.47	1.57	2.32	4.00	5.93	3.54	1.60	1.77	2.98	3.15	3.29	5.25	–	
16	BM	2.07	2.13	1.73	1.89	2.71	2.91	3.85	3.17	1.97	1.84	3.91	2.97	3.22	3.82	4.40	–
		MIN	CST	QA	CA	MPN	AD	BSL	MTO	SOL	CSE	CT	PM	C	LD	RP	(BM)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)

Table 4.1 *Mean average similarities ratings*

(Dimension II) directions on Figure 4.5. The occupational titles located at the extremes form the contrasts, and discontinuities are indicated by the gaps.

	Contrast			
	Positive Pole	*vs*	Negative Pole	Discontinuity
Dim. I	(CA, QA, CSE, CS)		(BSL, RP, LD)	MPN and CST
Dim. II	(MPN, PM, AD, MIN, CST)		(BSL, RP, QA, CA, BM)	C and CST

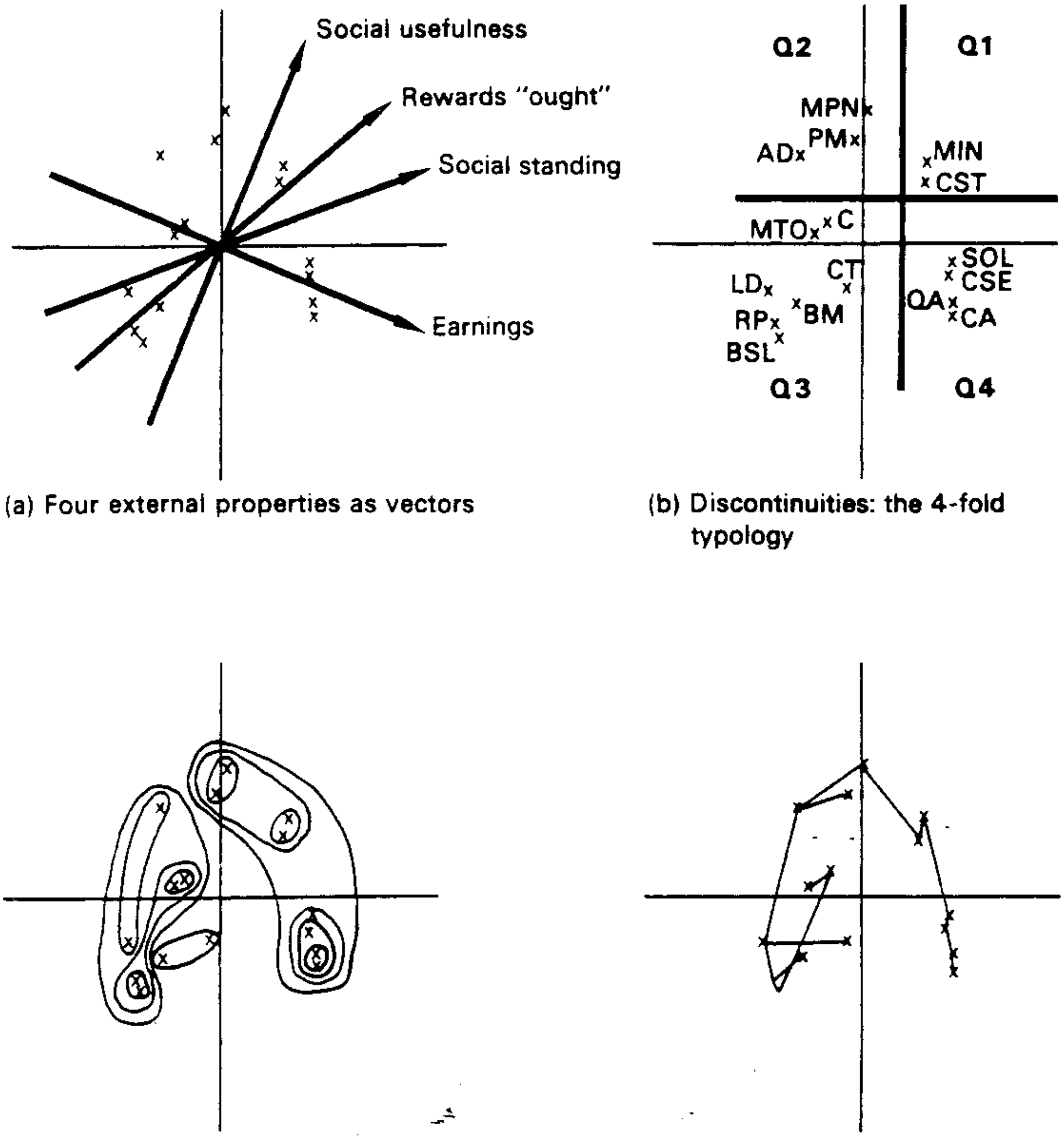
The terms used by the *subjects* to describe the contrast involved in Dimension I were retrieved by looking at what they said about the pairs concerned, e.g. Accountant *vs* Labourer; Accountant *vs* Porter; Civil Servant *vs* Porter; Solicitor *vs* Lorry Driver. The concepts they employed included 'qualifications', 'skills required', 'education', and we, as researchers, decided that the common core to the descriptions made the tag 'educational qualifications' a reasonable one. Of course, other connotations such as status and income were also present and other labels could just as easily have been chosen. Dimension II was named as 'service orientation' by a similar process. Note that, in the instance, we used subjects' descriptions of the contrast as a fundamental resource, but decided upon the final identifying label ourselves.

(ii) *External properties*
In the study, we treated occupational judgment as involving the analytically separable components of cognition and evaluation, the former operationalised in terms of similarity judgments and the latter in terms of the first three characteristics listed earlier. The subjects' ratings were then averaged, thus providing four 'external properties'.

Both point-distance location and vector representations were used (by means of the PREFMAP III and IV models). Without any doubt the vector representation gave a better fit to the data for each property, and these are presented in Figure 4.6a. An interesting point is that the most explicitly evaluative property, 'social usefulness', is independent of (at right angles to) estimated earnings: on average, people judge the pay for an occupation to be unassociated with its social worth. Moreover, the 'good' end of the vectors—whether of social utility, status or earnings—all point towards the professional side of the configuration.

Turning now to the differential regional density of the points, it is worth noting that, if the largest single gap on each spanning dimension is marked as a dividing line, the four quadrants each include the two occupations chosen in the original design to be examples of a four-fold typology combining 'educational requirements' and 'people-orientation'. (These occupations are emphasized in Figure 4.6b.) In effect, the subjects view the occupations very similarly to the researchers, which is a useful bonus.

The diameter HCS clustering of the averaged similarities data is presented in contour form in Figure 4.6c, up to the final two levels. The pattern is clear with either HCS: the major divide is between the left and right hand sides the manual and the professional occupations, with the Barman and Commercial Traveller largely unassociated.



(a) Four external properties as vectors

(b) Discontinuities: the 4-fold typology

(c) Hierarchical clustering

(d) Minimum spanning tree

Figure 4.6 *Various interpretations of single configuration*

The dense clusterings occur round the two professional groups (people- and non-people oriented) and within the skilled and unskilled manual groups. It is very obvious that the clustering follows very closely the horseshoe sequence shown in Figure 4.5 and not any dimensional direction.

(iii) *Graphic analysis*
Finally, graphic analysis was used to detect local structure and it was at this point that the horseshoe pattern became very evident. When mapping the lines into the configuration in sequence from the highest similarity down, it became very obvious that the pattern was that of a linked set of clusters, rather than a simple sequence of occupations. In addition to mapping the top quartile of similarities, the minimum spanning tree (MST) was also constructed. The purpose of constructing a MST (Prim 1957) is to connect the points by a network (or tree) of *connected* links which have the smallest *overall* distance. The MST is illustrated in Figure 4.6d. Once again, the horseshoe sequence is very apparent as the first ten links join up the

points from the lower right-hand corner *in sequence* round to the lower left-hand corner.

(iv) *Final interpretation*

But what does the horseshoe mean? The clue to this came from looking at the HCS and matching clusters of points at each level with what the subjects said about each cluster. Thus Actuary and Accountant were regularly described as 'concerned with figures'; when Solicitor and Civil Servant are added as 'instrumental occupations' or 'managing', and when joined with the Minister and Teacher ('teaching') and the Policeman and Nurse ('custodial'), the entire group are repeatedly called 'professions'. It thus became clear that different descriptions (predicates) existed at different levels of generality. The range of generality of each of the most common predicates was obtained by finding out to which occupations a given predicate typically applied. This was looked at and mapped into the horseshoe sequence, with the length of the arrow indicating the range of generality (see Figure 3.6 in Coxon and Jones 1978a). It then became clear that:

> The predicates repeatedly change as one moves along the 'horseshoe', making it very difficult to interpret the map as only involving a single contrast or dimension, even one as general as 'social status'. Yet there is a sequence, at least in the sense that predicates tend to develop and rarely appear outside their 'typical' range of applicability ... In moving from cluster to cluster along the sequence there are certainly many correspondences but equally, common features drop out and others re-appear.
>
> (ibid, p. 92)

Thus, while there is a very clear sequence present in the configuration—a highly non-linear one, not to be confused with a dimension—it turns out not to be a sequence of continuous meaning. Rather, it is a family of resemblances where characteristics or properties apply at different segments so that the sequence is more like a chain of associations or what Wittgenstein (1958, section 60) describes as a 'family resemblances' theory of meaning.

However abbreviated, this example makes it clear that different methods of interpretation often provide convergent evidence about both local and global aspects of the structure or configurations which, with a little imagination, can be invaluable in aiding interpretation.

7.5 Preference Mapping (PREFMAP I and II)

The Carroll-Chang (Carroll 1972, p. 114 et seq.) preference mapping models form a hierarchy of models, akin in many ways to PINDIS. They differ principally in that PREFMAP (PM) is designed primarily for *external* scaling (where the user provides a stimulus configuration) and the input data consist of a rectangular matrix of ratings or rankings of p stimuli given by N subjects. The purpose is to map *each subject* into the stimulus space in the form of an ideal point (PM Phases I–III) or as a vector (PM Phase IV) according to a hierarchy of increasingly complex models. (The transformation may be linear/metric or quasi non-metric/ordinal according to the user's choice.) We have already encountered earlier in this book the two simplest models: the *simple distance model* (Phase III) (5.3.3.1 and 6.2.4) and the *vector model* (Phase IV) (5.3.2 and 6.2.2) and so the focus here will be on the more complex models—the rotated and weighted distance model (Phase I, akin to P2 in PINDIS) and the weighted distance model (Phase II, akin to P1 and INDSCAL). But it will be helpful to begin by summarising the full range of the PREFMAP models.

7.5.1 The PREFMAP hierarchy of models

The basic notion of all the models in PREFMAP is that an individual's preference ranking is a function of the distance separating her point of maximum preference (ideal point) and the stimulus points. It is the way in which the distances are defined which differentiates the models.† (PREFMAP can also be seen as extending and generalising Coombs' unfolding model, and as showing that the vector model of preference is a special case of unfolding.)

Briefly, the hierarchy of models is as follows:

I *General Unfolding Model (PM1)*

This is the most general model. Each subject is viewed as having a specific, most-

*When a two-dimensional PINDIS analysis is performed, the results are much the same, but even more marked. Admissible transformations account for 75% of variation in the 'internal' PINDIS analysis and for 14% variation when related to the 'external' INDSCAL configuration. In neither case do the dimensional models add more than a derisory amount (2% in both cases), but the increase due to simple vector weighting is more impressive (18%, and 57% in the external case). But the 'unscrambling' is different in detail: the machine tool operator is still relocated more than any other occupation, but the actuary and accountant are also seen to be relatively unstable in their positioning.

†Technical details of the models are given in Carroll (1972) and in Coxon and Jones (1979, p. 106 et seq.), and a detailed exposition of the computational procedure and output details is given in van Schuur (1977).

preferred ideal point in the space, rotating the axes of the similarity space to his own reference dimensions, and then attaching an evaluative weight to each of them.

II *Weighted Unfolding Model (PM2)*
Subjects are assumed to have an ideal point and to share the same set of reference dimensions (i.e. no individual rotation), but to evaluate or weight the dimensions differentially.

III *Simple Unfolding Model (PM3)*
Subjects are assumed to share the same set of reference dimensions *and* to give the same weighting to them. However, subjects do still differ in terms of where their ideal points are located in the space. This is the external analysis analogue to Coombs' unfolding analysis.

IV *The Vector Model (PM4)*
Carroll (1972b) has shown this to be a special case of Coombs' unfolding, when an ideal point is located far from the origin of the space.* It is for this reason that it features in the hierarchy, since all the other models are straightforward *distance* models. In the vector model, by contrast, subjects are represented as a vector (or line directed toward the region of greatest preference), and a preference ranking is interpreted as the order of the projections of the stimuli points on this line. (The internal analogue of this model is MDPREF.)

One particularly valuable aspect of the hierarchical nature of the PREFMAP models is that it is possible to test whether a more complex model explains significantly more variation in the data than a simpler one. In this way, the model which makes the most parsimonious set of assumptions can be chosen. Moreover, as in PINDIS, there is no reason why one particular level or model should be thought to apply to all subjects—it might well be, for instance, that whilst the simple unfolding model applied to most subjects, the data of the remaining subjects might be far better explained by assuming that they simply differentially weight the dimensions of the space.

All the PREFMAP models make two basic assumptions:

(a) *A common similarity space is assumed to apply to all the subjects included in a* PREFMAP *analysis.* If it turns out that this does not hold (because, for example, a previous INDSCAL analysis of the similarities led to the conclusion that subjects were divided between distinct 'points of view'), then a separate PREFMAP analysis should be run for each subset, using the group space for each subset as the input configuration.

(b) *An individual's preference values for a set of stimuli are assumed to be linearly (or in the non-metric version, monotonically) related to the distance between her ideal*

*Intuitively this can best be seen in terms of isopreference contours (see Figure 7.12). For the basic distance model, all stimuli at a given distance from a subject's ideal point form a circle (isopreference contour) in 2-space, a sphere in 3-space, and so forth. For the vector model all stimuli at a given distance along a subject's vector form a line (projection) in 2-space. If an ideal point is taken further and further out from the origin, the circular isopreference contours come closer and closer to being a line in the vicinity of the stimulus points. See Carroll (1972).

point and the stimulus points. This assumption is only made for the first three models (which are distance models), and is not made for the vector model.

To give these ideas some substance, let us return to the example where a sample of respondents had been asked to assess the general similarity of a set of Irish politicians, and then been asked to rank them in order of personal preference. Let us also assume that the INDSCAL analysis of their similarity data indicated that they had fairly similar perceptions of the politicians and that the two main differentiating axes were left/right orientation and Republican/Unionist. The INDSCAL group space configuration could then be used as the 'independently established' cognitive space for input PREFMAP, and the focus of interest would now shift to explaining differences in the preference rankings of politicians which the respondents gave.

The simplest *vector model* (Phase IV) of PREFMAP assumes (as in MDPREF) that the subjects collapse the multidimensional stimulus space into one dimension (or line) representing the order of preference. Individual differences in preference are expressed by the differing directions which the vectors have in the common space. In this example, it is conceivable that some subjects simply prefer politicians who combine radicalness with support for Irish Republicanism (or conservatism with union with Britain), whilst others evaluate a politician solely in terms of how left-wing she is, or how much she supports Irish Republicanism.

The *simple unfolding (distance) model* (Phase III) assumes, by contrast, that each subject has one most preferred point in the group space, and that this serves as a reference point for evaluating the politicians, according to how close they are to her ideal point. The *weighted unfolding model* (Phase II) assumes that subjects differ considerably in the value they attach to the dimensions of the space—figuratively, that they pay attention to how highly they value or weight each dimension before they decide how close a politician is to their ideal. The effect of this is to 'pull' a politician closer to the subject's ideal point than would be the case in the simple unfolding model, if the politician occupies a high position on a dimension which the subject highly values.

The *general unfolding model* (Phase I) drops the assumption that subjects' evaluations refer to the same fixed set of dimensions. Instead, it allows them to structure the space as they wish by providing their own reference axes (so long as the dimensions they choose are not correlated) and *then* it allows them differentially to evaluate these 'private dimensions'. The effect of this is that subjects can be allowed to place high evaluation on *combinations* of the original dimensions. For instance, if the original axes were rotated anticlockwise through 45°, subjects who chiefly prefer politicians who combine socialism *and* republicanism (but who still differentiate socialist-Unionists from conservative-Republicans) could easily be accommodated.

Turning now to the hypothetical results of a PREFMAP analysis, it might turn out that, on average, the simple unfolding model (II) held best—that is, most subjects evaluated the politicians in terms of highly salient characteristics (dimensions) which were evaluated in the same way, and only differed substantially in what their positions were on the dimensions. But the data of a minority of subjects might be much better explained by assuming, *in addition*, that they valued highly politicians who were socialist-Republicans, cared not at all for conservative-Unionists but still made some (but relatively little) differentiation between socialist-Unionists and conservative-Republicans.

7.5.1.1 PREFMAP *Phases I and II*
Phase I: General unfolding model
Subjects are permitted

(i) to rotate the reference dimensions of the space, and
(ii) *then* to weight them differentially.

In Carroll's terms:

> We allow distinct individuals additional freedom in choosing a set of 'reference axes' . . . and then to weight differentially the dimensions defined by this rotated reference frame, in addition to being permitted an idiosyncratic ideal point.
> (Carroll 1972, p. 120)

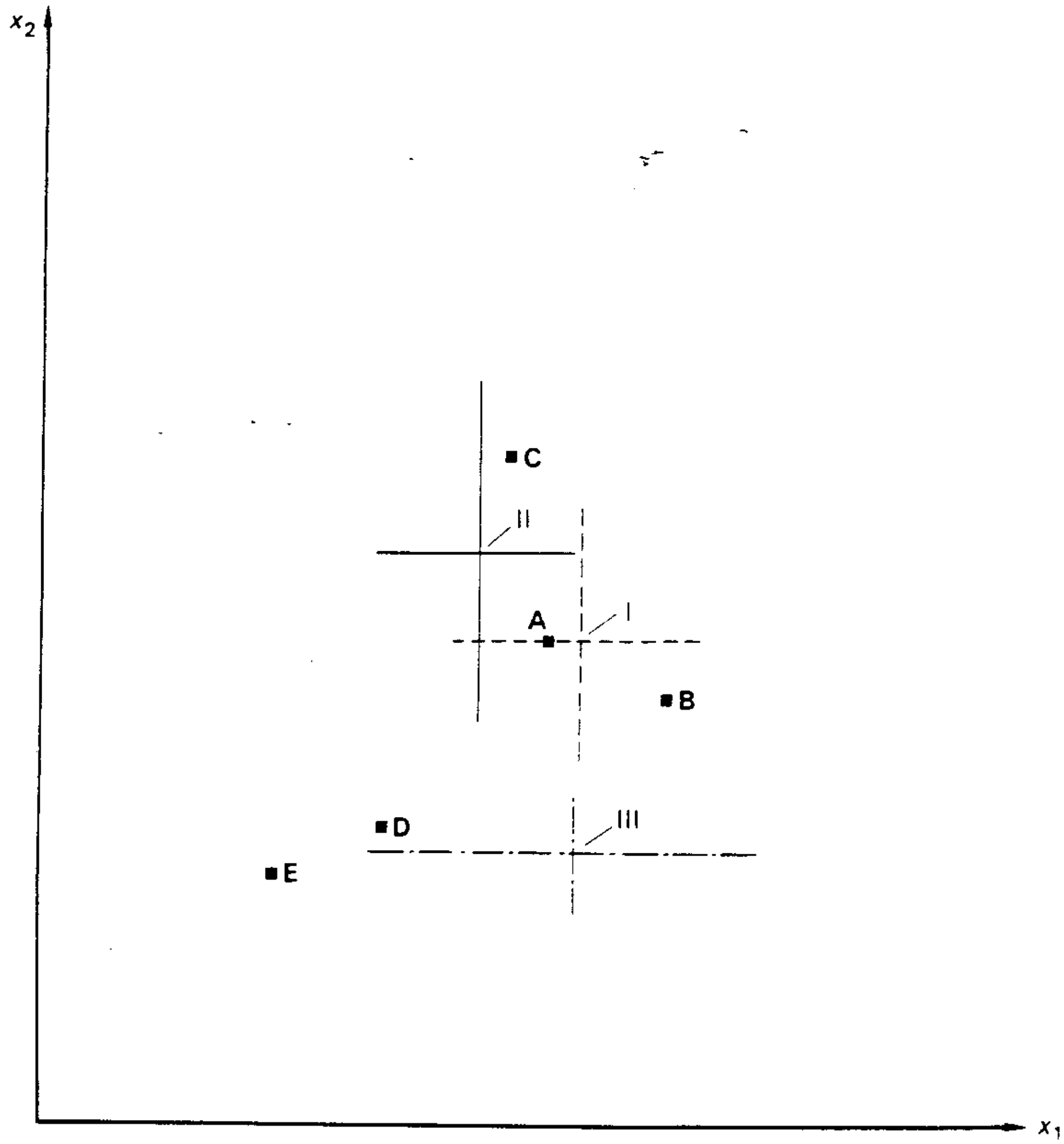


Figure 7.11 *PREFMAP, Phase II (weighted unfolding): subject ideal points and weighted dimensions*

A subject is assumed to apply his own orthogonal rotation to both the stimulus and ideal point co-ordinates, and *then* weight the rotated dimensions. If x^*_{ja} represents the *transformed* stimulus co-ordinates, and y^*_{ia} the transformed ideal points, then

$$s_{ij} = F_i(d^2_{ij})$$

where $d^2_{ij} = \sum_a w_{ia}(y^*_{ia} - x^*_{ja})^2$,

i.e. a Euclidean distance in an individually-rotated and weighted 'private space'.

Phase II: Weighted unfolding model
This is similar in many ways to the INDSCAL and PINDIS P1 model except that in this *preference* model, the weights are interpreted as reflecting the subjects's *evaluation* of the dimension when making an overall preference judgment. To continue the Irish example—two subjects might well entirely agree about, say, the left-right orientation of the politicians. To one subject this may override all other considerations when it came to choosing a candidate, whilst to another it might be considered entirely irrelevant compared to the politician's position on Republicanism.

In this model, a subject is assumed to apply an evaluative weight w_{ia} to each dimension, so that

$$s_{ij} = F_i(d^2_{ij})$$

where now $d^2_{ij} = \sum w_{ia}(y_{ia} - x_{ja})^2$,

i.e. a Euclidean distance in a weighted 'private space'.

The weighted unfolding model (PM2) is illustrated in Figure 7.11 (and see also Figure 7.12). In Figure 7.11 the joint space includes five stimulus points (A to E) and the location of the ideal points of three subjects (I, II and III). In this model, the subjects differ in terms of the location of their ideal points (akin to PINDIS idiosyncratic perspective model (P4)†), and they also attach different evaluative weight or salience to the dimensions. In Figure 7.11 the differential weights are denoted in the form of the arms of a cross. In the case of I, the weights for the dimensions are equal; for II the weight of dimension II is greater than that for dimension I, and for III the reverse is true. Note that in this model the individual axes are all oriented in the same direction, parallel to the reference axes and are therefore directly comparable. 'Private spaces' for each individual can be produced, if so desired, by stretching and shrinking the reference axes. A more convenient representation is illustrated in Figure 7.12.

7.5.2 The *a priori* stimulus space

Since the PREFMAP models are designed to be external in form, the user must normally supply an *a priori* configuration, and similar issues arise as in the case of PINDIS: what may reasonably be used as an *a priori* configuration? How, if at all,

†Compared to the PINDIS hierarchy, PM2 is a hybrid model. PM2 is similar to P4 in the sense of allowing differing points of view, but is similar to P1 in allowing differential weighting of axes.

does the configuration change at different levels?

The source of the *a priori* configuration can be considered under three heads: (i) a previous scaling, (ii) a theoretical or rational configuration and (iii) an 'internally-generated' configuration.

A previous scaling
The configuration may be the result of a previous MDS scaling analysis of similarities data for the same set of stimuli, probably obtained from the same subjects that provided the preference data. This is probably the most common instance in social science applications.

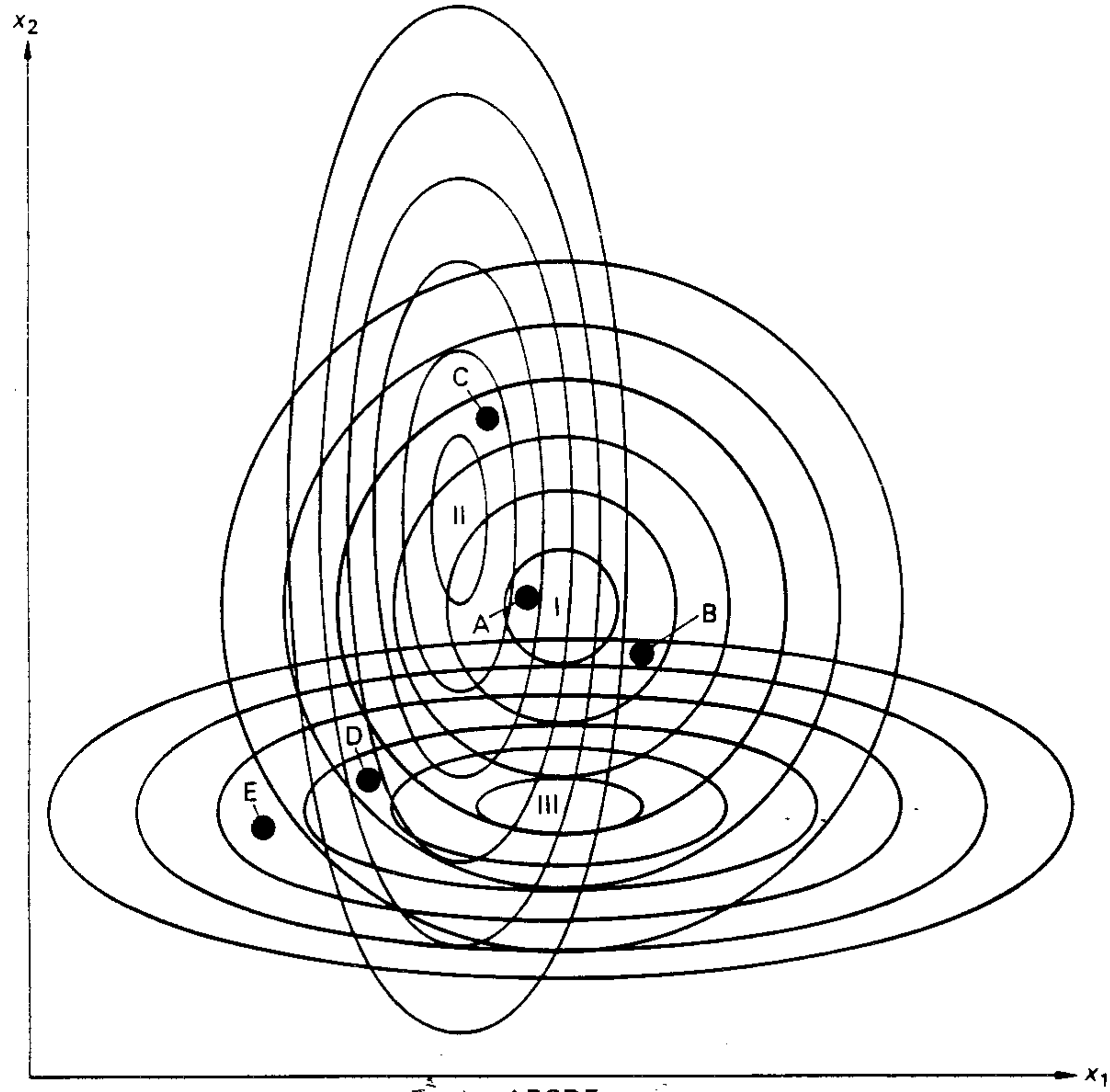
When replicating a study it can be useful to see how well the data from one's own study will fit the configuration obtained by the original investigator.

A theoretical or rational configuration
By a 'rational configuration' is meant one which incorporates the actual characteristics of the stimuli as dimensions of the *a priori* space. This occurs typically in psychophysical applications and in cases where the stimuli are well-defined compositions. Examples include the now infamous 'hypothetical cups of tea' data (collected by Wish and analysed in the Carroll 1972 paper) whose defining characteristics were the *hotness* and the *sweetness* of the brew, and in the Delbeke-Bollen family composition data (cf. Coxon 1974 and section 6.2.2) whose characteristics were the *number of sons* and *number of daughters* making up a family. Other instances might include a facet analysis of the stimuli, the geographical location of stimuli, or some other theoretically expected or physically underlying configuration.

Internally generated configuration
The option also exists in PREFMAP for a stimulus configuration to be constructed from the *same* preference rankings data as are used in the analysis. (It is implemented in the MDS(X) version using the INITIAL parameter.) Such a 'quasi-internal' configuration is constructed by forming the minor (stimulus × stimulus) product moment matrix from the data, (which may have been transformed in some user-specified way*), and then performing a classic basic scaling (see Appendix A5.2) to yield the configuration.

In the MDS(X) version of PREFMAP, the user also has the freedom to keep the stimulus space configuration fixed throughout all phases of analysis, or to allow the program to modify it to obtain a better fit to the data by the use of the KEEP parameter. If the first option is chosen, the original stimulus configuration remains unchanged throughout the analysis. In the second case, the form of modification is slightly different, depending upon the level of the model (phase) at which the user starts analysis (see Carroll and Chang 1967, p. 10 et seq. and Carroll 1972, p. 134 et seq.).

*Options exist within the INITIAL parameter for double centring the product matrix (removing row and column means and replacing the overall mean); removing row means; standardising rows, and for removing row and column effects (see *User Manual*, PREFMAP Report, 2.3.1). If the first option is chosen and the data then analysed under the linear version of PM3, the resulting analysis turns out to be equivalent to an internal metric unfolding analysis. On this and related issues, see Carroll 1970, 1971, Carroll and Chang 1970, and Schonemann 1970.



I – ABCDE
II – CADBE
III – DEBAC
(Reproduced with permission from Carroll 1972)

Figure 7.12 *Weighted unfolding model (PREFMAP Phase II): isopreference contours for three individuals*

7.5.3 Preference and 'anti preference'

In the distance (unfolding) model of preference, it is assumed that a subject's preference is single peaked and symmetric—that she has one point of maximum preference in the stimulus space and that preference decreases systematically in all directions (along each dimension). One way of representing this is to define a set of contours (at arbitrary but equal intervals) centred upon the ideal point. Each contour represents points at an equal distance from the subject's ideal point, and hence all points will be equally preferred. In preference analysis, these contours are referred to as 'isopreference contours'. In the case of the simple distance model (PM3), where each dimension is equally weighted, these contours will in the two-dimensional case be circular, and in the weighted distance model (PM2) they will be ellipses the length of whose axes will equal the value of the evaluative weight. This is illustrated in Figure 7.12, which presents the isopreference contours for the three subjects of Figure 7.11. Note that subject I has equal weights (and hence circular preference contours) whereas the others have unequal, and hence elliptical contours. Once the preference contours are drawn in, it becomes straightforward to reproduce a subject's preference ranking by following the relevant contours. Thus, for subject II, C is within the second ellipse from the ideal point. A within the third, D within the fifth, B just beyond the sixth and E further beyond it—yielding the I-scale: CADBE.

In the case of the rotated and weighted model (PM1) the ellipses are oriented in different directions, but the same logic applies. In the case of the vector model (PM4) the isopreference contours are straight lines, rather than circles or ellipses, since all points which project onto the same point along the preference vector, wherever they be located in the space, will be equally preferred.

The PREFMAP models specifically allow for negative evaluative weights. How are they to be interpreted? Carroll argues that it is simply a matter of changing the form of the preference function in the distance models, from being single-*peaked* and symmetric to having a single valley (one single point of *minimum* preference) and symmetric. In this case, the contours will decrease towards the subject's 'pessimal', 'anti-ideal' or least preferred point. (Perhaps the simplest example is the preference for the heat of liquids: many people dislike lukewarm tea most of all, but increasingly prefer both hot and iced tea.) It will sometimes happen that a subject's point will include a mixture of negative and positive weights; this situation has been explored in Carroll (1972, pp. 121–3), who shows that in the two-dimensional case the function will be saddle-shaped. Nonetheless, it is often difficult to give meaning to such mixed-sign combinations and some care is needed in their interpretation.*

7.5.4 Goodness of fit

As a nested set of models, it is possible to use analysis of variance statistics as an indicator of how much better one model does than another (or how much more variation one explains) than another. At the end of a PREFMAP run a table is printed containing correlations, F ratios for each subject by each phase/model and between pairs of phases/models, together with root mean squares (RMS) for each phase. The correlations are between the squared distances of the model and the preference data values (or corresponding disparities in the case of the quasi-nonmetric option having been chosen).

The average RMS values usually give an indication of the level of the hierarchy of models which is generally most appropriate, i.e. that level where the marginal increase in RMS values for more complex models is very small. The subject-phase correlations can indicate whether any phase fits a subject's stated preferences, and the extent to which one phase fits better than another. In conjunction, these measures enable the user to decide upon the most appropriate model (phase) of PREFMAP for representing the data, and they also allow subsets of subjects to be allocated to different levels.

Since each model is a special case of the one immediately above it, it is justifiable

*Davison (1976) has proposed and recently implemented (Davison 1980) a variant of preference mapping which allows the evaluative weights to be constrained to non-negativity or non-positivity.

to test for goodness of fit of a model to the data by the usual ANOVA procedures, and test for significant increases in variation explained by means of an F-test between models, although the models are clearly not independent.

Most users enter the PREFMAP program at a higher phase/model and drop down to a lower phase. Carroll (1972, p. 135) argues that the solution for the 'average subject' of the highest phase entered should form the basis for the succeeding phases.

Thus, if S-PHASE (1) and E-PHASE (4) are chosen then the full set of models will be applied to the data. This will mean:

(i) After Phase I (PM1), the rotated reference axes of the average subject form the 'canonical reference frame' for Phase II (PM2).

(ii) The subject's weights and ideal points are fitted in *this* reference frame in Phase II (PM2), rather than in the original input frame.

(iii) In a similar manner, the average subject's dimensional weights are applied to the rotated configuration for analysis in Phase III (PM3).

These operations are not always desirable, and the remedy is either to return to the original input configuration at each phase, by setting KEEP (1) or to enter at a lower phase. For instance:

to keep the *orientation* of the original configuration, enter at PM2
to keep the dimensions unweighted, enter at PM3.

The parameters of the model are estimated through quadratic and linear regression (for the metric version) and subsequent monotone regression (for the non-metric version). The former allows statistical testing, on the assumption that the data and fitted values are linearly related, but the latter only permits approximate statistical tests and they should be used with caution.

In many applications of PREFMAP, the main differences seem usually to occur between the simple distance (unfolding, P3) model and the simple vector (P4) models, with few convincing examples of the necesity of invoking more complex models except for particular subjects.

7.5.5 Uses of PREFMAP

The PREFMAP program can be used in three principal ways:

(i) to map individual ratings or rankings into an independently obtained external configuration (external joint mapping).

(ii) to first estimate the stimulus configuration from the subject's data, and then map the same rankings/ratings into it (quasi-internal mapping).

(iii) to use PREFMAP to extend two-mode scaling to include additional subjects (extending two-mode solutions).

7.5.5.1 External joint mapping

The first is undoubtedly the most common use and the one for which PREFMAP was originally designed. In most cases, the researcher has obtained both pairwise similarities data and preference rankings or ratings from the same subjects. The similarities data will typically have been scaled by INDSCAL, or averaged over all subjects and scaled by MINISSA, and the resulting stimulus configuration (group stimulus space in the case of INDSCAL) is then input as the external configuration.

7.5.5.2 Quasi-internal mapping

So far there are no published examples of quasi-internal preference mapping. However, as mentioned above (see Carroll and Chang 1971), choice of the metric option FIT (0) with a double-centred configuration generated from the preference data GENERATE (0) provides an optimal solution for an internal *metric* unfolding analysis which, given the frequent instability of the non-metric unfolding analysis (implemented by MINIRSA), may well be a desirable option to choose.

7.5.5.3 Extending two-mode solutions (large data sets)

Users often want to scale data sets where the number of subjects is too large for stated program limits. In these circumstances PREFMAP may be used to effect an increase in the number of stimuli or subjects. A reasonable strategy is to proceed as follows:

(i) Sample the maximum permitted number of subjects and include the relevant data in a preliminary run of scaling programs to produce a 'core scaling'.

(ii) Take the output configuration and treat it as an 'external' configuration for PREFMAP analysis.

(iii) Select the options within PREFMAP which match the model and transformation of the original scaling.

(iv) Include the remaining data as input 'preference' data for the PREFMAP run, which will then estimate the positioning of the new subject points (or vectors) within the existing configuration.

Although this procedure is most appropriately used for extending the number of subjects for two-way, two-mode scaling models (such as MDPREF and MINIRSA), it may also be used for basic one-mode scaling models (such as SSA and MRSCAL) if the additional data consist, for each new point, of a set of relative distance estimates between the new point and the points already in the fixed configuration. Kruskal (1972b, pp. 5–6) shows how he used essentially this procedure when faced with scaling a data matrix of 10,000 rows (computer malfunctions) by 657 columns (diagnostic tests) to get a configuration of 10,000 points. In essence, subsets of a manageable size were first used to get a rough estimate of the structure and dimensionality of the data. Then a core set of points (in Kruskal's case, 20) is chosen in such a way that they are as well spaced in the configuration as possible. Finally, the remaining points are positioned into the configuration with respect to the core points, by metric or non-metric scaling, as implemented by PREFMAP. This and other ways of scaling large data sets are discussed in Golledge et al. (1981).

Care should be taken to ensure that the transformation is correctly specified (by the parameter FIT) and that the correct model is chosen—S-PHASE (3) for distance models and S-PHASE (4) for vector models.

7.6 Interrelations of INDSCAL, PINDIS and PREFMAP Models

The more complex models discussed in this chapter have a family resemblance.

INDSCAL, PINDIS and PREFMAP each consists of a set of models of increasing complexity, and the type of complexity is also alike, including distance models involving differential dimensional weighting and rotation, and one or more vector models. The relationship can be portrayed as follows:

Model	Program/Series (a) INDSCAL	(b) PINDIS	(c) PREFMAP
Rotated, Weighted Distance	(IDIOSCAL)	P2	PM1
Weighted Distance	INDSCAL	P2	PM2
(Simple Distance)	SSA/MRSCAL	P0	PM3
Vector	—	P3, P4	PM4

The parallelism is not exact, and not all of the programs exist in the MDS(X) package but, arrayed in this way, certain common elements become clear:

(a) The INDSCAL series is the most basic, and take their input from one or more (dis)similarity matrices. The most complex model, IDIOSCAL (IDIosyncratic Orientation SCALing) is a fairly predictable generalisation of INDSCAL where each subject is thought of as carrying out an idiosyncratic rotation of the group stimulus space axes, followed by a weighting of those axes. Unfortunately, as a program it has a number of sub-optimal characteristics, and far from convincing examples of its use have been published. The basic references are Carroll and Wish (1973, pp. 90–2 and 1974, pp. 440–1).

The basic metric and non-metric models, MRSCAL and MINISSA, are obviously special cases of INDSCAL—when each subject has identical sets of weights. There also exists in the original MINI series (and elsewhere) variants of the basic vector ('factor analysis') model, (see, for example, Lingoes et al. 1979, pp. 268–9 (SSA-III) and pp. 307–9 (MINI-NFA)). The widespread availability of (metric) factor analysis programs and of the closely related classic scaling procedure in MRSCAL make its inclusion in the MDS(X) series superfluous. In its non-metric form it has rarely been used, since product moment correlations are (inversely) monotonic with distances (see Appendix A2.1), making it a somewhat redundant model.

(b) The PINDIS series distance models correspond closely to their INDSCAL analogues, except that they take individual configurations as input. The PINDIS series also includes the vector series and the hybrid P5 (double-weighting) model.

(c) The PREFMAP series differ from the other two series in taking rectangular (row-conditional) data as basic input, and representing each row element by an ideal point (distance models) or an ideal vector. Otherwise, the models correspond to their analogues in the other series.

It should also be clear by now that the form of transformation chosen is relatively unimportant compared to the form of the model. Indeed, it is rather ironic that after the 'non-metric revolution' has been accomplished it turns out that in many cases the linear (metric) assumption is a good (and considerably less costly) approximation to the monotonic one. That said, it is as well to err on the side of caution and allow the extensive family of monotonic functions to suggest what more regular transformation might be more appropriate.