MAPPING THE MIND WITH BROKEN THEODOLITES:
Contributions to Multidimensional Scaling methodology, with special application to Triadic data, and the Sorting and Hierarchical Sorting Methods.

A dissertation submitted in
partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Human Development Studies
at Massey University
by

D. L. Bimler

Massey University
May 1995

BIMNER

Massey University Library

Thesis Copyright Form

Title of thesis: Mapping the Mind with Broken Theodolites

Please delete section not applicable.

(1) (a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.

(b) ~~I do not wish my thesis to be made available to readers without my written consent for .... months.~~

(2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

(b) ~~I do not wish my thesis, or a copy, to be sent to another institution without my written consent for .... months.~~

(3) (a) I agree that my thesis may be copied for Library use.

~~(b) I do not wish my thesis to be copied for Library use for .... months.~~

Signed [signature]

Date 17/5/95

* *

For Readers Only.

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME and ADDRESS **DATE**

ABSTRACT

This thesis focuses on the psychological applications of Multidimensional Scaling (MDS) theory and methodology. The results are investigated of treating certain kinds of dissimilarity data (triadic data, to begin with) as comparisons between dissimilarities. This is a familiar idea but many of its implications are unexplored.

First, when data are available from more than one subject, it becomes possible to apply models of individual variation, in non-metric form. The Weighted Euclidean (or INDSCAL) model is the one used most often in this thesis, but the more general IDIOSCAL model is used to investigate individual differences in the case of colour vision. The data sets need not be complete. This is important when the size of the stimulus set means that there are too many comparisons for a single subject to respond to them all.

Second, Maximum Likelihood Estimation (MLE) becomes a straightforward generalisation of the standard hill-descent algorithm for minimising Stress.

Third, data collected with the sorting and hierarchical sorting methods can also be regarded as dissimilarity comparisons. The convenience of the sorting method and the lesser demands it makes on subjects when the number of stimuli is large have led to its widespread use, but the best way of analysing such data is uncertain. A 'reconstructed dyad' analysis is described and shown to be better than the usual co-occurence approach in a number of examples in which evidence about the true perceptual or conceptual space is available independently.

Finally, when the data are interpreted as dissimilarity comparisons, an interactive method of scaling large stimulus sets becomes possible, in which one selectively acquires incomplete data, concentrating on comparisons which are expected to contain most information about the configuration. This approach has been applied twice, with the stimuli being simple synthesised sounds in one example, and complex natural sounds (canine heartbeats) in the second, working well in both cases. The potential applications for training people to recognise sounds are briefly considered. Some possibilities for future research arising from this work are described.

ACKNOWLEDGEMENTS

Some of the ideas involved in this thesis where presented to the IVth International Workshop on Infant Cry Research, Munich, July 1992. I am grateful to the organisers of the Workshop, whose financial help made it possible for me to be there; also to Prof. Dr. Dietmar Todt, of the Institut für Verhaltensbiologie, Freie Universität Berlin, who provided hospitality and accommodation in Berlin, and to Peep Stalmeier, of the Nijmegen Institute for Cognition and Information, who made me equally welcome in Nijmegen.

The goals of my research were largely theoretical, but without the generosity of a number of people who made available data they had collected, for me to re-analyse, in addition to my own data, it is doubtful whether I could have made what progress I have. The people to whom I am indebted include Peep Stalmeier (colour triads), Willem van der Kloot (power-strategy F-sorts), Rich Faldowski and Mark Hollins (texture F-sorts), Stephen Morley (pain F-sorts), and the ESRC Data Archive at the University of Essex. As well as information, they provided interested encouragement, as did other researchers who could not help with the former.

In particular, the data collected by Emma Barraclough of Massey University – partly for her own research, on colour vision and market research, but partly on my request – have been invaluable. I can only hope that I have been as much of assistance to her as she has to me. The informants (judges, subjects) themselves deserve to be thanked many times over.

The loan of a computer by Dr Roger Jones, of the Research School Social Sciences, Australian National University, meant that a two-month exile in Canberra was not wasted.

I should thank my principal supervisor, John Kirkland, who has shown unfailing faith that I knew what I was doing, in the face of overwhelming evidence to the contrary. There are numerous other people I should thank, and I can only hope that they will understand that the omissions are from the pressure of time, and not from lack of appreciation of their help. I must also acknowledge the cogent and constructive criticism of journal reviewers.

To Penny, who laughs at my jokes.

TABLE OF CONTENTS

PARTIAL LIST OF SYMBOLS

N	number of elements (items, stimuli)
i,j,k,l	usually element indices, $1 \leq i,j,k,l \leq N$
(E_i,E_j) or (i,j)	dyad consisting of the i-th and j-th elements
N_d	number of possible dyads. $N_d = N(N-1)/2$
$<E_i,E_j,E_k>$ or $<i,j,k>$	triad consisting of i-th, j-th and k-th elements
N_t	number of triads in a data set. For a complete set, $N_t = N(N-1)(N-2)/6$
P	number of dimensions
p	usually a dimension index, $1 \leq p \leq P$
X	a reconstructed configuration of N points (a single point in $(N\,P)$-dimensional configuration space)
x_i, x_{ip}	position of the i-th element in X
$\mathbf{X}$	configuration considered as a (N-by-P) matrix
Δ	matrix of dissimilarities
δ_{ij}	element of Δ. $\delta_{ij} > \delta_{kl}$ is equivalent to writing $(i,j) \gg (k,l)$.
D	matrix of reconstructed distances
d_{ij}	element of D
Đ	matrix of disparities ('pseudo-distances')
$ð_{ij}$	element of Đ
$\varepsilon_{ij,kl}$	distance comparison coefficient. $\varepsilon_{ij,kl} = 1$ if $\delta_{ij} > \delta_{kl}$, 0 otherwise.
M	the number of subjects
m	usually a subject index, $1 \leq m \leq M$
D_m, Δ_m	dissimilarities, reconstructed members for subject m (elements are $d_{m,ij}$, $\delta_{m,ij}$)
$\varepsilon_{m,ij,kl}$	distance comparison coefficient for subject m (comparing $\delta_{m,ij}$ and $\delta_{m,kl}$)
$\mathbf{W}$	(M-by-P) matrix of dimensional weights (saliences)
w_{mp}	element of $\mathbf{W}$
C_m	sorting data co-occurence matrix for subject m
$c_{m,ij}$	element of C_m
E	matrix of averaged co-occurences
e_{ij}	element of E. $e_{ij} = 1/M\ \Sigma c_{m,ij}$
f_{ij}	corrective force exerted between the i-th and j-th elements

1. INTRODUCTION

Within the experiments and theoretical work reported in this dissertation, three general directions can be discerned. In order to summarise them, I separate the three objectives, as follows:

1. Improving the analysis of certain widely-collected forms of psychological dissimilarity data, by extensions of multi-dimensional scaling techniques (MDS);

2. When performing a MDS analysis on a data set which combines data collected from several subjects, to recognise differences between subjects. Any contrasting of differences must be done without forfeiting the greater reliability given to the map by the larger data set.

3. Eliminating redundancy in the data, so as to ease the burden of data collection for subjects (or alternatively, to keep the burden the same while allowing a greater number of stimuli to be scaled).

A thorough discussion of MDS is not the province of this Introduction, but some explanatory remarks are necessary before expanding upon these objectives.

MDS makes it possible to construct a geometrical model of a restricted perceptual or semantic domain, using information about the dissimilarity of stimuli in that domain, in the experience of subjects. Stimuli are represented by points in the model; distances between points reflect, as accurately as possible, the experienced dissimilarity between stimuli.

An analogy is reconstructing a map of a country from a roadmap that has been damaged and oil-stained until the only part still legible is the table of inter-city travel distances. However, MDS models are not limited to two dimensions. The dimensionality is generally a compromise, with more dimensions providing a better reflection of the dissimilarities, at the expense of decreased simplicity.

Point-to-point distances are not the only geometrical metaphor for modelling dissimilarity or proximity data. Network structures, trees in particular, are one widely-used alternative [e.g. Arabie, 1991; Carroll, 1976; Pruzansky, Tversky & Carroll, 1982; de Soete, DeSarbo & Carroll, 1985] – though they are outside the scope of this dissertation. For more general definitions of 'distance' (involving more than two points), see Junge [1991].

In perceptual domains, examples of stimuli might be Munsell colour chips, or notes played by different musical instruments; examples of conceptual stimuli are "numbers (0 to 9)", or "occupations". Henceforth, the terms "stimulus", "element" and "item" will be used interchangeably, since MDS is abstract enough to make no distinctions between the varied natures of the raw material it's used on. Thus no attempt is made to analyse the actual stimuli, or to access whatever qualities that might be contained in them; they are treated as benchmarks, for a task which has parallels in surveying.

Many complex phenomena fail to lend themselves to analysis or quantification, but they can still be studied with MDS. A particular example of interest to developmental psychologists is baby cries. These complex sounds can be analysed into a variety of acoustic properties [Gustafson & Green, 1989; Michelsson, Raes, Thoden & Wasz-Hockert, 1982], but it is not obvious which of these correlates with the qualities that make one cry more aversive to listeners than another, or distinguishable from non-cry sounds [though see Zeskind, 1987; Zeskind & Marshall, 1986]. These qualities survive a certain amount of distortion, e.g. filtering by intervening walls. They could not serve their evolutionary function if they relied upon perfect acoustics.

Given a table of perceived similarities between pairs of these auditory stimuli, MDS would derive a map of "cry space". The map's dimensions (i.e. independent ways in which one cry can differ from another) could perhaps be interpreted in terms of acoustic measurements. Clusters of points would provide a taxonomy of cries and hint at distinct classes amongst them. The researcher would then be in the favourable position of studying cries using readily-obtainable equipment, which despite its cheapness has been fine-tuned over countless generations to respond to their salient properties: the human ear. There is no need to approach the cries with a battery of scales for measuring them, with the associated risk of prejudicing the research's outcome, since scales can omit important qualities, or over-represent irrelevant ones.

But are such similarities available?

Applications of MDS allow proximities to be obtained in several ways. They can be estimated indirectly, by "stimulus generalisation" methods. An illustration is provided by a classic example of the MDS technique: Rothkopf's statistics of how often one Morse Code symbol is mistaken for another, which Shepard used to create a two-dimensional map of "Morse Space" [Shepard, 1962], interpreting them as a measure of the symbols' proximity. Confusion data are obtained more easily if there is a set of labels or pigeon-holes, "right answers", with which subjects can categorise the stimuli. In market research, an indication of the similarity between products comes from how frequently consumers switch their preferences. However, these indirect methods limit the size of the sensory or semantic map, since items must resemble one another enough to be confused or exchanged (failing that, to increase confusion probabilities, it is necessary to degrade the observing conditions [e.g. Miller and Nicely, 1955] or restrict observation time [e.g. Killam, Lorton & Schubert, 1975; Plomp, Wagenaar & Mimpen, 1973]).

More directly, one simply asks subjects to rate the dissimilarities, e.g. on a scale of 1 to 10, or by making a mark somewhere along a line. However, the complexity which makes some stimuli into candidates for MDS research, also hampers evaluation of their dissimilarities. The human mind lacks the kind of internal yardstick required to assess distances between pairs of sounds, colours, smells, or concepts consistently. When presented with a series of cries, each pair displaces the previous pair, leaving no stable, calibrated context for making the assessments. It comes as no surprise that when one group of cry researchers asked their subjects to repeat their judgments, the inferred reliabilities (Pearson correlations between repeated cry pairs) ranged from 0.94 down to 0.29 – and that was after eliminating three subjects with consistencies between 0.23 and -0.11 [Green, Jones & Gustafson, 1987].

Often one is forced to combine data from several subjects, which makes the inconstancy of the yardstick worse. But when there are many items it is impractical for every pair to be assessed by a single subject. If, instead of using the Method of Sorting (with overlapping item sets so that each subject only sorted 90 items), Kraus, Schild and Hodge [1978] had elicited pairwise data in their study of 220 occupation terms, they would have needed 24 090 comparisons (or a significant proportion thereof). In a study on the vocabulary of pain, Verkes, van der Kloot and Van der Meij [1989] scaled 176 adjectives, again using the Method of Sorting, whereas 15 400 pairwise comparisons would have been necessary.

Generally such problems arise in applications of MDS within psychology. I ignore the many MDS applications in other fields, where data collection is straightforward and their analysis is untroublesome. One exception to this policy of neglect is the study of social groups: later chapters will call upon Struhsaker's observations of vervet monkey sleeping groups, and Sampson's observations of social distances within a monastery.

In these problematical situations, where we are in the position of surveyors unable to measure the distance from one benchmark to another, we resort to more roundabout approaches to data acquisition (examples are the triadic procedure, and sorting into groups), which deliver their results directly at a higher level of abstraction than a table of dissimilarities. To quote an apt description: "Every response from assessors in these methods can be [...] reached by a process of binary comparisons of dyads, where every comparison results only in a decision about which dyad spans the greater subjective distance" [MacRae, Howgate & Geelhoed, 1990, p. 697].

I concentrate for the moment on the triadic procedure. This involves presenting elements three at a time, accompanying each with questions equivalent to "Are I and J more alike than I and K, or less so?" Each *triad* provides its own context. It is analogous to a theodolite that is broken but can still inform the surveyor which of two landmarks is more remote from a reference point.

I will present an approach for performing MDS with triadic data, and demonstrate its advantages over the commonly-used "vote-counting" approach of converting the data into a table of dissimilarities.

The artifacts introduced by vote-counting into the final configuration are not necessarily serious, given a complete set of triads, which contains enough redundancy to mask the distortions. However, a complete data set is not always available. As the set of elements grows larger (greater than 12, say), it is preferable to take advantage of the redundancy to reduce the number of triads, which otherwise becomes prohibitively large, increasing as the third power of the element number. When it is not practical to confront each informant with a complete list of stimulus triads, I will look at the question of choosing the optimal incomplete list.

Whatever the experimental procedure, a subject's capacity to provide information is limited by factors such as fatigue and boredom. Unavoidably there is a ceiling on the number of points in the space under scrutiny, where the sparseness of an individual's data set forces one to pool data from several subjects. This raises the issue of how individual variations fit into the geometrical model: a question I have so far begged.

Individual variations cannot be ignored when the data have been collected via the Sorting procedure (or the related hierarchical sorting procedure, or in the form of ranked preferences). Scaling such data relies on some disagreement between subjects; complete unanimity results in degenerate configurations.

Individual variations can be treated as errors, and multiple subjects regarded as experimental replications, by assuming that the "space" one is mapping is held in common, while any differences from one replication to the next are brought about by random, non-systematic effects. However, some research has the identification of individual differences as its objective. One may be investigating colour blindness. Developmental psychologists are interested in how perceptual and semantic maps evolve with age and experience. In the baby-cry example, there is evidence that parental status and exposure to the stimuli affects one's "cry map". It has been suggested that an important aspect of parent-infant dynamics is the ability, not granted to everyone, to differentiate one kind of cry from another.

At the very least, a version of MDS tailored for processing triadic data should offer an equivalent to the INDSCAL option of standard MDS. I will focus on continuously-varying models of individual variation, ignoring the latent class model [Meulman & Verboon, 1993; Winsberg & de Soete, 1993] (in which "latent classes" or sub-populations are postulated; it is not initially known which class each subject is a member of; each class has a distinct configuration, shared by its members).

I will attempt to convince the reader that even in situations where the data sets are too incomplete to scale a single subject's data in isolation, such sets can be combined in a way that leaves open the possibilities of assessing individual variations, and making meaningful comparisons between subjects.

These same points apply to the sorting procedures. They similarly benefit from a version of MDS which by-passes the preliminary Procrustean conversion of the data into a dissimilarity table – a form of vote-counting. Being less redundant than triadic data, they are more vulnerable to vote-counting artifacts.

With these initial observations out of the way, I am in a position at last to outline the contents of the following chapters.

Chapter 2 will delineate MDS in greater detail, revisiting the pioneering work of Shepard, Guttman, Kruskal, *et al*. Johnson's approach, the most versatile, is the one I concentrate on. I argue that "vote-counting" data before scaling them is both undesirable and unnecessary.

Analyses of real data are introduced in Chapter 3, in which the triadic procedure is covered, as a way of giving concrete value to the theoretical points made previously. The examples include five applications of the triadic procedure to auditory stimuli: cries, in one example; synthesised sounds in the other four. MTRIAD, a program for analysing these data, makes its first appearance.

Chapter 4 will return the focus to the mathematical background. Generalisations are described which allow variations between subjects to be analysed, and which accommodate the possibility of non-Euclidean "spaces".

In Chapter 5, inspired by work by Ramsay [1977, 1978], Takane [1981] and Takane and Carroll [1981], I introduce Maximum Likelihood Estimation (MLE), implemented in a novel way. MLE has a number of advantages, helping, for instance, with the problem of degenerate solutions.

Chapter 6 will generalise the discussion to various forms of ranked data, of which triadic data are a special case; other special cases are conditional ranking and ranked preferences.

The cases of the sorting and hierarchical sorting procedures are interesting enough to devote a chapter to them, which is Chapter 7, the longest by far. Some of the stimulus sets used in the examples have already appeared in Chapter 6, providing independent 'maps' of stimulus spaces to corroborate the ones obtained from sorting data. The inadequacies of the standard methods for analysing sorting data have not detracted from the popularity of the procedure.

Unfinished business remains from Chapter 3. Chapter 8 will look at variations of the triadic procedure which collect incomplete data, in order to circumvent the limitations which arise from the third-power proliferation of triads. Among these methods is an interactive one, modelled on ISO [Young, Null & Sarle, 1978].

Chapter 9 will draw conclusions and point out possible directions for future study.

This thesis began as a far more circumscribed project. On a first encounter with triadic data (for cries), I felt that the existing procedure for scaling them was distortion-prone and wasteful of information, and that an improved method was possible which would better reflect the time and concentration invested by the subjects who had provided them. Unaware that Roskam had pointed out the same flaws in vote-counting in 1970, I set about tailoring a program to the data. It subsequently turned out to re-state Johnson's 1973 algorithm.

As well as the advantage of its versatility with different data formats, Johnson's formulation of MDS seems to be appealingly intuitive, leading to its frequent rediscovery. See, for instance, the series of papers by Schneider and his co-workers.

The work reported here may seem dated. But despite recent theoretical progress in MDS, I believe that the "classic" approaches still repay close attention, with much extension and elaboration remaining to be done before they are completely mined-out.

The practitioners of MDS are mainly concerned with wringing every drop of meaning out of finite data sets. If the literature is any guide, MDS is generally applied in situations which do not warrant enormously sophisticated mathematics or models of mental functioning. With this in mind, I offer small improvements to the collection and analysis of triadic and sorting data, in the belief that the number of potential users makes them worth reporting.

Reflecting a computational bias, I prefer to spell the nested summations of the equations out in full, even when concise matrix forms are available.

Current or anticipated applications of this work include:

- a contribution to the long-standing debate on the dimensionality of facial expressions;
- market research on Paulownia wood involving large numbers of wood-block stimuli;
- the taxonomy of complex sounds (cries; heartbeats), with implications for classifying them by acoustic criteria only;
- feedback for students being trained to diagnose such sounds;
- scaling adjectives used in self-descriptions of pain.

Many of these will surface in later chapters, demonstrating different facets of the theory, amid the inevitable Monte Carlo simulations and the borrowed and synthetic data sets.

2. MULTIDIMENSIONAL SCALING

"Multidimensional scaling" is not a particularly helpful name for the techniques I describe in this chapter. The name derives from the older tradition of one-dimensional scaling: the process of assigning scalar values to items, i.e. sequencing them along a single axis, on the basis of information about their relative positions (Thurstone's Case V comparisons).

Borg and Lingoes [1987] point out that the term "MDS" has the misleading connotation that the dimensions are meaningful, in that items should lie along each axis in a sequence, as in the one-dimensional case. To shift the emphasis away from premature interpretation of the axes, they prefer "multidimensional similarity structure analysis". Similarly motivated, Guttman and Lingoes had previously coined the phrase "smallest space analysis" to describe their MINISSA suite of programs [Guttman, 1968], but familiarity counts for more than clarity, and MDS remains the term in widest circulation.

The common goal of all MDS techniques has already been introduced: in lieu of analysing a set of stimuli, to represent them as simple points in a geometric model, arranged so as to uncover the structure (if any) of the similarities experienced between pairs of stimuli. The goal can be reached by a variety of paths. I will not attempt to survey the field thoroughly (a Sisyphean task, given the rate of progress, with new advances published on a monthly basis). This chapter dwells only on the details relevant within a limited, rather revivalist perspective. For a broader coverage of MDS, see Schiffman *et al* [1981], or Borg and Lingoes [1987].

For the moment, with a loss of generality which can be corrected later, I use the Pythagorean formula for "distance" to model similarities.

Let X be the configuration, in a P-dimensional Euclidean space. Given a set of N elements, $\{E_1, E_2, \ldots E_N\}$, each is represented by a point

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, \ldots x_{iP}).$$

Keeping $i < j$ to avoid duplication, the number of dyads or pairs of stimuli, (E_i, E_j), is $N_d = N(N-1)/2$. The distance separating $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is

$$d_{ij} = \left(\sum_{p=1}^{P}(x_{ip} - x_{jp})^2\right)^{1/2} \tag{2.1}$$

Distances are invariant under arbitrary rotations of the configuration. This is one reason for caution when interpreting the dimensions. These d_{ij} are entries in an N-by-N matrix D.

In the simplest case, data takes the form of a matrix of dissimilarity ratings, Δ. The matrix entries δ_{ij} are initially assumed to be proportional to d_{ij}, apart from random errors:

$$\delta_{ij} = \Phi(d_{ij}) + \text{(random error term)} \qquad \text{where } \Phi(d_{ij}) = C_1\, d_{ij}.$$

I find it fruitful to imagine building a simple analog device for scaling such data, by taking N nodes and linking them with N_d springs, where the spring linking the i-th and j-th nodes has an equilibrium length δ_{ij}. When this contraption is held in some initial configuration $X^{(0)}$ and then released, I imagine it flexing furiously, springs stretching and contracting and dragging the nodes this way and that towards some optimal arrangement which minimises the strain on them. To start with they overshoot but the reverberations slowly die away as equilibrium emerges.

Computer simulation of this spring model is an intuitively-appealing way of performing MDS. One iteration will not be enough: when a corrected configuration is produced on the basis of the clashes between the distances of the initial configuration and the ideal spring length, there will be a degree of overshoot, taking some springs farther from equilibrium, but eventually an arrangement results which is not necessarily free of strain, but any residual spring forces acting each node are in balance.

The iterative approach is superfluous: for this simple case of ratio data (i.e. distances in the model are assumed to be proportional to dissimilarities), an eigenvalue solution is available [Young & Householder, 1938]. The case of interval data is only slightly more difficult: the unknown constant C_2 in

$$\Phi(d_{ij}) = C_1\, d_{ij} + C_2$$

is assigned the smallest value necessary to make the triangle inequality true [Torgerson, 1952].

These are "metric methods". The case of ordinal data, where nothing is assumed about the psychophysical function $\Phi(x)$ linking dissimilarities and stimulus-space distances except that it is monotonic, requires non-metric methods, which recover $\Phi(x)$ concurrently with the configuration. The Introduction has already mentioned some examples of ordinal-level data, collected indirectly. As well as judged similarity and confusion rates, other possibilities include response latency time, error rates in paired associates tests, the degree of generalisation of a conditioned response, galvanic skin response, time taken to sort multiple copies of two stimuli into separate groups (constrained classification tasks), and cross-modal interference (the Stroop effect).

Non-metric methods have become the norm; "non-metric" is taken for granted when MDS is mentioned without further qualification (an exception being INDSCAL and related software).

The point about ordinal-level dissimilarities or proximities is that the actual values acquired are irrelevant, and only their relative values matter: whether $\delta_{ij} < \delta_{kl}$, $\delta_{ij} > \delta_{kl}$, or $\delta_{ij} = \delta_{kl}$. For the purposes of analysis, dissimilarity (or proximity) ratings can be reduced to inequalities.

I write $(E_i,E_j) \ll (E_k,E_l)$ if $\delta_{ij} < \delta_{kl}$, i.e. if stimulus E_i is judged more similar to E_j than E_k is to E_l.

In some experiments a rank order is obtained directly, bypassing any assignment of values δ_{ij}, by asking subjects to rank all N_d dyads in order of increasing dissimilarity [e.g. Shepard & Chipman, 1970]. This is a full rank order; every dyad's position relative to every other one is known:

$$(E_{i1},E_{j1}) \ll (E_{i2},E_{j2}) \ll (E_{i3},E_{j3}) \ll \ldots$$

or simply $$(i1,j2) \ll (i2,j2) \ll (i3,j3) \ll \ldots \qquad (2.2)$$

labelling the stimuli with their indices, sacrificing clarity for the sake of concision.

Shepard [1962] argued that for large enough N, ranked data serves as well as metric ratings for the purpose of reconstructing the configuration.

With modifications, the spring model accommodates ranked data. Instead of the (i,j)-th spring exerting a corrective force on i and j, dependent on how far its length deviates from an ideal length δ_{ij}, each spring has a ***relative*** ideal position in the ranking order. One must imagine the (i,j)-th spring consulting with all others, accumulating contributions to the force it exerts, expansive contributions for every spring longer than it although before it in the rank order (2.2), and contractive ones for every spring shorter than it although putatively longer (according to (2.2)). The magnitude of each contribution is proportional to the corresponding distance discrepancy.

I call this the "smart spring" model.

Expressing the ranking order relationships (2.2) in the form of coefficients $\varepsilon_{ij,kl}$, similar to Guttman's "signature" coefficients [1968]:

$$\varepsilon_{ij,kl} = \begin{cases} 1 \text{ if } (i,j) \gg (k,l) \\ 0 \text{ otherwise,} \end{cases} \tag{2.3}$$

and giving the springs a spring constant κ, then the spring device has total potential energy

$$\begin{aligned} \text{P.E.} &= \kappa \sum_{(i,j) \gg (k,l)} \mathrm{H}(d_{kl} - d_{ij})\,(d_{kl} - d_{ij})^2 \\ &= \kappa \sum_{(i,j)} \sum_{(k,l)} \varepsilon_{ij,kl}\, \mathrm{H}(d_{kl} - d_{ij})\,(d_{kl} - d_{ij})^2 \\ &= \sum_{(i,j)} \mathrm{E}_{ij}, \qquad \text{where } \mathrm{E}_{ij} = \kappa \sum_{(k,l)} \varepsilon_{ij,kl}\, \mathrm{H}(d_{kl} - d_{ij})\,(d_{kl} - d_{ij})^2 \end{aligned} \tag{2.4}$$

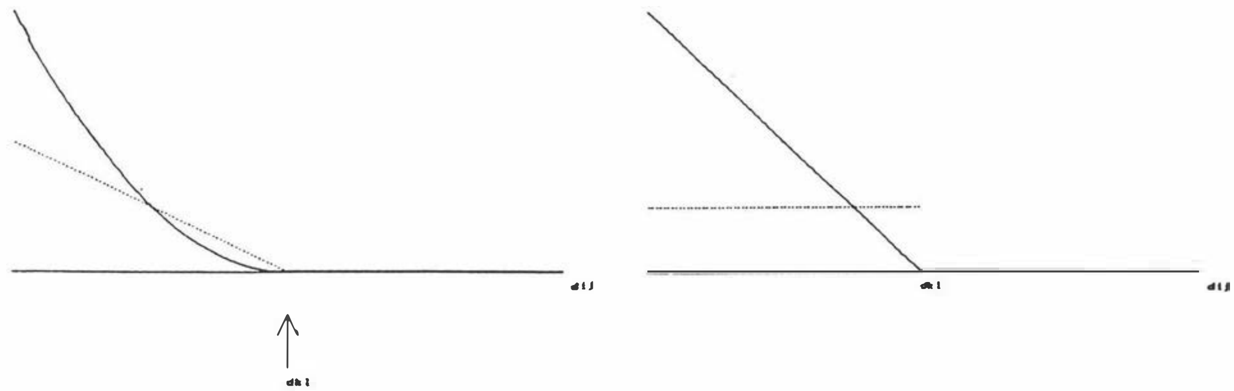
Here H(x) is the Heaviside step function, $\mathrm{H}(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x \le 0 \end{cases}$

When summing over dyads, it is understood that $i < j$, to avoid duplicating contributions when implementing the summations as computer programs.

Then
$$\begin{aligned} \partial \mathrm{E}_{ij} / \partial d_{kl} &= 2\,\kappa\, \varepsilon_{ij,kl}\, \mathrm{H}(d_{kl} - d_{ij})\,(d_{kl} - d_{ij}) \\ &= 2\,\kappa\, \varepsilon_{ij,kl}\, \Theta(d_{kl} - d_{ij}) \end{aligned}$$

where Θ(x) is the ramp function, $\Theta(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \le 0. \end{cases}$

Figure 2.1. Contributions to potential energy (solid line) and spring force (dotted line) of (i,j)-th dyad from comparison with (k,l)-th dyad, for two sorts of spring, (a, left) and (b, right)



The magnitude of the (i,j)-th spring force f_{ij} is derived by summing these contributions:

$$f_{ij} = 2\,\kappa \sum_{(k,l)} \varepsilon_{ij,kl}\, \Theta(d_{kl} - d_{ij})$$

Its direction comes from the fact that the force on i and j is exerted parallel to the spring. The total force on x_i is the vector sum of such forces; its components are

$$f_{ip} = -2\,\kappa \sum_{j} (x_{ip} - x_{jp}) \,/\, d_{ij} \sum_{(k,l)} \varepsilon_{ij,kl}\, \Theta(d_{kl} - d_{ij}) \qquad (2.5)$$

Young [1975] provides a similar geometrical derivation. The springs in this case are obedient to Hooke's Law. There are other possibilities. A case can be made for linear contributions to the potential energy, as in figure 2.1(b). Convergence seems faster, with the oscillations dying away more rapidly, and a single large discrepancy is less likely to dominate several small discrepancies.

$$E_{ij} = \kappa \sum_{(k,l)} \varepsilon_{ij,kl}\, H(d_{kl} - d_{ij})\, (d_{kl} - d_{ij}),$$

and $$f_{ip} = -\kappa \sum_{j} (x_{ip} - x_{jp}) \,/\, d_{ij} \sum_{(k,l)} \varepsilon_{ij,kl}\, H(d_{kl} - d_{ij}) \qquad (2.6)$$

Certain caveats must be made. Though I have treated rating data as reducible to rank orders, the two are different cases in the theory of data. When a subject sorts proximities into ascending order, as in the earlier example of Shepard and Chipman, I assume that each of the implicit comparisons between dyads is actually made; it becomes legitimate to speculate about the chance of a given comparison result being erroneous: this is part of the Maximum Likelihood approach to MDS (Thurstonian pairwise data provide the comparisons explicitly). One cannot make this assumption for ratings.

There is information in ratings which rankings lack. Thus the second caveat: the discussion so far has ignored "quasi-metric" forms of MDS, which treat the data at a level intermediate between interval and ordinal. In these, the function modelling the proximities belongs to a restricted family (polynomials, or sums of exponential functions, or splines, for instance), instead of allowing an unrestricted monotone function. Fewer degrees of freedom are involved. The stronger function assumptions may be based on psychophysical considerations, or they may simply impose the restrictions of non-abruptness or convexity. Some algorithms for MDS allow the user to specify the admissible transformations.

The values δ_{ij} are retained in these quasi-nonmetric approaches. They are not applicable when the inequalities (2.2) are all that is known, as in the rank-ordered data sets already mentioned, and in the more general forms of inequality data covered by subsequent chapters (triadic, conditionally rank ordered, etc.)[1].

Fully rank-ordered data (whether obtained directly, or derived from dissimilarities) is only one special case among many. More general, less structured forms of dyad comparison do not necessarily possess its two important properties: completeness, and transitivity. Putative inequalities $(i,j) \gg (k,l)$ obtained through a Thurstonian procedure of pairwise comparisons [e.g. Bissett and Schneider, 1992] are not guaranteed to be transitive. Triadic data are incomplete: two dyads (i,j) and (k,l) are only compared and their relationship known if they belong to the same triad, sharing an element in common. With three comparisons per triad, and $N\,(N\text{-}1)\,(N\text{-}2)\,/\,6$ triads, the maximum number of inequalities is $N\,(N\text{-}1)\,(N\text{-}2)\,/\,2$, instead of $N_d\,(N_d\text{-}1)\,/\,2$. The same is true for data obtained through the conditional rank-order procedure. A converse possibility is that (i,j) may be compared against (k,l) more than once. So even if such data are consistent, the dissimilarities cannot be arranged in a linear, monotonic order (2.2); a complex network of relationships is all that's possible. This is *partial rank order*.

How does this affect scaling?

---

[1] Some dyad comparisons may disagree between one experimental replication and the next. Given enough replications, it is possible to estimate the (unmeasured) δ_{ij} values by applying the Law of Comparative Judgment to such disagreements: e.g. Torgerson's analysis of triadic data, or the approach used by De Soete and Winsberg [1993] to reconstruct the preference function from pairwise preference comparisons.

The inequalities of the data cannot be treated as absolute constraints, for as we have seen, there is no guarantee that they are consistent, even before any distortions of the distances required to cram them into a low-dimensional space. The optimal solution is one where the inter-element distances contradict the data inequalities to the smallest possible extent.

Define a configuration space, having $(N \cdot P)$ dimensions, in which a given configuration X can be represented by a single point, the coordinates of the elements being listed in vector form rather than as a matrix:

$$X = (x_{11}, x_{12}, \dots x_{1P}, x_{21}, x_{22}, \dots x_{2P}, \dots, x_{N1}, x_{N2}, \dots x_{NP})$$

Some possible points X violate many inequalities; others violate few. Kruskal [1964] had the inspiration of quantifying the degree of that violation with a function which he called STRESS(X). [also Guttman, 1968]. Subsequently, alternative definitions have been put forward. It will be worth examining several, to check their flexibility, i.e. how far each one relies on the data following a standard format.

For now, I use S(X) as a generic label for an unspecified badness-of-fit measure. Other possible names for it are "height", or "potential energy", imagining the configuration space as an $(N \cdot P)$-dimensional landscape ranging from valleys of low data violation up to mountains where the violations are maximal. At each X there is a gradient, $\nabla S(X)$. Its components $\partial S(X) / \partial x_{ip}$ can be considered as components of forces $\partial S(X) / \partial x_i$ which push the i-th point in directions which reduce S(X) by lengthening or shortening the various d_{ij}. In this light, an obvious way of seeking out the deepest valley – not the fastest – is a hill-descent algorithm. Given $X^{(t)}$, a non-optimal configuration, the next approximation

$$\begin{aligned} X^{(t+1)} &= X^{(t)} + \Delta X^{(t)} \\ &= X^{(t)} - (\text{step size})\ \nabla S(X^{(t)}) \end{aligned} \tag{2.7}$$

This is continued until S($X^{(t)}$) stabilises. There is no guarantee that following the slope downhill has led to a global minimum. There may be many local minima, each with its own basin of attraction, its own "watershed" in the configuration-space landscape.

The step size should be large enough to reach the minimum in a reasonable number of iterations, and to have a good chance of jumping across the smaller "watersheds" of local minima, without continually overshooting in violent oscillations. Following Kruskal's lead, MTRIAD (our implementation of MDS) uses an adjustable step size which is increased when consecutive vectors $\Delta X^{(t)}$ point in the same direction ($\Delta X^{(t)} \cdot \Delta X^{(t-1)} > 0$) and decreased in the case of an overshoot, when $\Delta X^{(t)}$ backtracks on the previous vector ($\Delta X^{(t)} \cdot \Delta X^{(t-1)} < 0$).

Another important issue is the choice of starting configuration $X^{(0)}$. MTRIAD estimates its $X^{(0)}$ by performing metric MDS (Principal Coordinates Analysis) on a table of dissimilarities, themselves estimated by "vote-counting" the data.

This algorithm is not the fastest. A priority has been to keep MTRIAD within the size limitations of an IBM PC running DOS. The exact details are not coupled tightly to other aspects of MDS, such as the particular definition of S(X), so there is the option of installing a faster algorithm, later (such as estimating step size with the conjugate gradient method, or using second derivatives of the Stress in the hill descent [de Leeuw, 1988; Ramsay, 1978; Takane 1978a]).

Some Alternative Definitions.

A desirable feature for S(X) is scale-invariance. Arbitrary expansions of X (as well as translations and rotations) should not affect S(X). Normally, the lowest possible value – for an X meeting all constraints – is 0. Conversely, the greatest possible value should be 1.

Johnson [1972] defines a lack-of-fit function θ^2:

$$\theta^2 = \sum_{(i,j) \neq (k,l)} \delta_{ij,kl} (d_{kl}^2 - d_{ij}^2)^2 \Big/ \sum_{(i,j) \neq (k,l)} (d_{kl}^2 - d_{ij}^2)^2$$

where his $\delta_{ij,kl} = \begin{cases} 1 & \text{if } \mathrm{sign}(d_{ij} - d_{kl}) \neq \mathrm{sign}(\delta_{ij} - \delta_{kl}) \\ 0 & \text{otherwise.} \end{cases}$

Clearly the values of the dissimilarity measurements δ_{ij} do not contribute to θ^2, only their rank order, so binary pairwise comparisons provide a valid form of input (this is implied by Johnson's term for the algorithm, "Pairwise non-metric MDS". A similar possibility is implicit in Guttman's [1968] explanation of Smallest Space Analysis). If dyads have not been compared, the corresponding terms should be removed from the denominator as well as the numerator (and equally, if pairs of dyads have been compared more than once, the replications can be incorporated into the equation with the use of another summation). Bissett and Schneider [1992] expand on this point.

Changing the order of the summation, and eliminating terms duplicated in numerator and denominator:

$$\theta^2 = \sum \varepsilon_{ij,kl} \mathrm{H}(d_{kl} - d_{ij}) (d_{kl}^2 - d_{ij}^2)^2 \;/\; \sum \varepsilon_{ij,kl} (d_{kl}^2 - d_{ij}^2)^2 \qquad (2.8)$$

The d^2 terms in (2.4) are a source of concern: they give undue weight to comparisons involving larger dissimilarities. Since there is nothing vital about squaring the distances, only a matter of algebraic convenience, allowing θ^2 to be expressed and solved in elegant matrix form, I have no qualms about substituting linear terms into (2.4), producing a slightly different function

$$\mathrm{S} = \text{raw Stress} \;/ \sum_{(i,j)\ (k,l)} \varepsilon_{ij,kl} (d_{kl} - d_{ij})^2 \qquad (2.9)$$

$$\text{where raw Stress} = \sum_{(i,j)\ (k,l)} \varepsilon_{ij,kl} \mathrm{H}(d_{kl} - d_{ij}) (d_{kl} - d_{ij})^2 \qquad (2.10)$$

Approximate the denominator as constant (making this Guttman's "soft squeeze").

Then $\partial(\text{raw Stress}) / \partial d_{ij} \quad = -2 \sum \varepsilon_{ij,kl} \mathrm{H}(d_k - d_{ij}) (d_{kl} - d_{ij}) \quad = -2 \sum \varepsilon_{ij,kl} \Theta(d_{kl} - d_{ij})$

The definition of $d_{ij}^2 = \sum_{p}^{P} (x_{ip} - x_{ip})^2$ gives $\partial d_{ij} / \partial x_{ip} = (x_{ip} - x_{jp}) / d_{ij}$

$$\partial(\text{raw Stress}) / \partial x_{ip} = -2 \sum_{(i,j)} (x_{ip} - x_{jp})/d_{ij} \sum_{(k,l)} \varepsilon_{ij,kl} \Theta(d_{kl} - d_{ij}) \qquad (2.11)$$

This is familiar from the spring model. They are different metaphors for the same algorithm. Instead of 'hill-descent', one might equally well describe the iterative process as 'relaxation'. The relaxation metaphor is a reminder that an efficient order for performing the calculations is to iterate first over the known comparisons, accumulating a matrix of spring tensions; then iterate over springs, resolving each tension into components of the force on its end-point elements, and accumulating those force components.

KYST, MDSCAL and POLYCON use a Stress function expressed in terms of how extensively the reconstructed d_{ij} need changing to bring them into the desired rank order [Schiffman *et al*, 1981]. In an intermediate step which Kruskal calls "monotone regression", the three algorithms all calculate "pseudo-distances" or *disparities* $ð_{ij}$ (or Ð as a matrix), which are correctly rank-ordered, but which are otherwise as close as possible to the d_{ij}. Plotting $ð_{ij}$ against d_{ij}, deviations from a straight line through the origin should be as small as possible. This is achieved by adjusting d_{ij} to maximise the coefficient of congruence, or equivalently, to minimise a raw Stress defined as the least-squares difference between D and Ð:

$$\text{normalised STRESS}^2 = \sum_{(i,j)}(d_{ij} - ð_{ij})^2 \;/\; \sum_{(i,j)} d_{ij}^2$$

The monotone regression is roundabout. Element dyads are first listed in order of their putative dissimilarity. When the distances for a sequence of dyads within that list violate the desired order, order is restored by replacing them with their mean value. For a given dyad (i,j) immediately to the right of (k,l), d_{ij} should be greater than d_{kl}. If $d_{ij} < d_{kl}$, both dyads are assigned disparities $ð_{ij} = ð_{kl} = (d_{ij} + d_{kl}) / 2$. This averaging of neighbouring dyad distances is extended until it produces a monotonic sequence of $ð_{ij}$ values. Kruskal shows that this definition minimises $\Sigma(d_{ij} - ð_{ij})^2$.

If we tinkered with the Stress, trying different potentials, consistency demands a corresponding change to what quality between distances and disparities is minimised in the definition of the latter. Using a ramp function instead of a quadratic potential energy would require us to replace order-violating distances with their mode, rather than their arithmetic mean, thereby minimising $\Sigma|d_{ij} - ð_{ij}|$.

For fully rank-ordered data and large N, this algorithm gains a speed advantage over Johnson's, since each dyad is compared only with immediately adjacent neighbours in the list, not with all other dyads. The advantage is lost during modification of the algorithm to handle general binary comparison data. A network of inequalities replaces the one-dimensional list, and instead of being limited to two, the number of adjacent neighbours to be checked against d_{ij} for consistency becomes open-ended. The modification requires sprawling data structures, nested iteration loops, and programming infelicities too numerous to contemplate.

Roskam [1970] has described a restricted generalisation of the Kruskal Stress in his program MINITRI which brings triadic data under the aegis of MDS. Chapter 3 will return to this.

The Stress used by Lingoes and Guttman in their MINISSA software is similar, except in providing an alternative definition of $ð_{ij}$, the "rank image". The rank-image sequence is made monotone by permuting rather than averaging the d_{ij}: if $(i,j) \gg (k,l)$, but $d_{ij} < d_{kl}$, then $ð_{ij} = d_{kl}$, $ð_{kl} = d_{ij}$.

Computationally this is more efficient than Kruskal's definition of disparities, and Guttman argues that it results in faster convergence. This approach has been found to less liable to entrapment in local minima. However, the concept of the rank image does not extend to partial rank order data, so the Lingoes-Guttman Stress does not generalise.

To wind up this non-exhaustive presentation of alternative Stresses, let us consider the ALSCAL program, one of the Alternating Least Squares suite [Takane, Young & de Leeuw, 1977; Schiffman *et al*, 1981]. This is not a hill-descent algorithm but disparities still play a part in it. For $X^{(t)}$, Ð is calculated, and subjected to *metric* MDS, producing $X^{(t+1)}$. Whether the algorithm generalises to non-transitive or partial rank orders depends on whether Kruskal's definition of disparities is used, or the Guttman and Lingoes rank-image definition.

Young proves that this process converges, minimising a Stress in the process: for a single subject,

$$\mathrm{SSTRESS}(1)^2 = \sum_{(i,j)}(d_{ij}^2 - ð_{ij}^2)^2 \Big/ \sum_{(i,j)} d_{ij}^4$$

This is a case where the discovery of the Stress function being minimised came after the algorithm minimising it. Note the squared distances. Weinberg and Menil [1993] found that these gave undue prominence to order violations involving large distances, affecting ALSCAL's performance to the extent that metric MDS performed better even when the simulated data were non-metric.

What conclusion is to be drawn from this look "under the hood"? The chief lesson is that there is no intrinsic limitation in the central concepts of MDS, that only permits data in the form of a table of proximities or dissimilarities to be scaled. One can construct an algorithm similar to Johnson's Pairwise Comparisons scaling, and capable of handling less structured sets of dyad comparisons, not necessarily complete or transitive. Other algorithms can be modified for the general case as well (though less easily).

The next chapter will go into more detail about MTRIAD, the computer implementation of this algorithm, looking at it in the context of triadic data. But before that, there remains the usual treatment of triads to consider, to whit, simply converting the data into a table of dissimilarities δ_{ij} (a process Coombs [1964] calls "decomposition"), suitable for processing by any of the standard software suites. We have seen that these dissimilarities are intermediate values, which the software promptly discards, retaining only their rank-order relationships, in the hope that these are not too badly distorted a version of the relationships in the original data.

"Vote-counting" is a commonly-used method of decomposition, especially with triads [e.g. Burton & Nerlove, 1976; Levelt, van der Geer & Plomp, 1966; Plomp, 1970]. The vote-count estimate for δ_{ij} is simply the number of comparisons between (i,j) and other triads (k,l) in which (i,j) is judged to be less similar:

$$vc_{ij} = \sum_{(k,l)} \varepsilon_{ij,kl}$$

In the case of unbalanced incomplete data sets, where the number of comparisons each dyad takes part in is not constant, the word "number" in this definition should be "proportion":

$$vc_{ij} = \sum_{(k,l)} \varepsilon_{ij,kl} \Big/ \sum_{(k,l)}(\varepsilon_{ij,kl} + \varepsilon_{kl,ij}) \qquad (2.12)$$

This approach seems plausible enough. vc_{ij} is high when i and j are distant. But situations arise in clustered configurations whereby a particular (i,j) gains a high vote count despite being close together, through comparisons against even closer pairs. Burton and Nerlove [1974] give glaring examples of the distorted configurations this can produce, when the data are incomplete.

Furthermore, with a maximum value for vc_{ij} of $2(N-1)$ in the triadic case, cases will arise of multiple candidates for the position of most separated dyad, all acquiring the same vote count and equal rank in the rank ordering.

In general, vote counting introduces new relationships, purporting (i,j) to be more distant or closer or equal to (k,l), when the subject has made no judgment because the two dyads were not compared, and their actual relationship is unknown. This is additional to ranking judgments made by the subject but over-ridden in the vote-counted version. Roskam [1970] called attention to these problems, with Monte Carlo simulations to back up his case.

Perhaps there are less Procrustean procedures for decomposing data into dissimilarities. Coombs [1964] describes a tabular "triangular analysis" technique (so called because it involves permuting the N_d-by-N_d half-matrix), intended to minimise the number of raw-data judgments over-ridden. Its usefulness is restricted by its reliance on an error rate lower than those commonly encountered in practice: intransitivities in the rank order are assumed to be rare enough that they occur in isolation, without forming compounded intransitive cycles, which are harder to correct. Deciding where to break such cycles often comes down to an arbitrary choice.

More complex indices of similarity obtained by summing comparisons have been proposed. There is no need to discuss them in detail since they all fail a criterion put forward by Torgerson: "The relative distances amongst any three stimuli, for example, should remain the same regardless of what other stimuli are included in the experiment." See Torgerson [1958] for details; Gladstones [1962b] compares two of these vote-count variants against Torgerson's analysis in practice.

Despite this, I have looked at a second-order vote-count dissimilarity index:

$$vc^{(2)}_{ij} = \sum_{(k,l)} \varepsilon_{ij,kl}\; vc_{kl} \;= \sum_{(k,l)} \sum_{(m,n)} \varepsilon_{ij,kl}\, \varepsilon_{kl,mn}$$

The summation proceeds in a wider context. Dyads (m,n) do not need to be compared with (i,j) directly in the experiment – they can contribute to $vc^{(2)}_{ij}$ through intermediate dyads (k,l) which were compared to both. The effects of the limited range of values are less, since the maximum attainable $vc^{(2)}_{ij}$ is $O(N^2)$.

Monte Carlo simulated experiments show this function to be an improved estimator of d_{ij}. However, Torgerson's observation still applies.

The basic problem remains: decisions about ordinalising the dyads should be made in the broadest possible context. Whenever error is greater than zero, contradictions and intransitivities appear in the data, and the information needed to resolve them is distributed through the entire data set. Distilling it (to use Shepard's choice of words) takes more than a formula; deciding whether to over-ride a dyad comparison to allow the configuration to fit other data depends on basic scaling details such as the dimensionality of the space being used. The only efficient way to recover the contextual information we seek is to embed the items in a P-dimensional space, and use the scaled distances d_{ij} to weight the comparisons for vote-count estimates. In other words, the ideal preliminary for MDS is MDS. This may sound paradoxical, yet it is the principle behind ALSCAL.

Later chapters will pursue the shortcomings of vote-counting further.

3. TRIADS

A set of triadic data can be expressed as a list of inequalities between dissimilarities, of the form

$$\delta_{ij} > \delta_{kl}, \qquad \text{i.e. } (i,j) \gg (k,l)$$

where i, j, k, l range over a list of N perceptual elements (stimuli), and (i,j) must have one element in common with (k,l) for the two dyads to be compared.

A number of ways of obtaining these inequalities are available. First described was the "method of triadic combinations" [Klingberg, 1941; Richardson, 1938]. As originally formulated by Richardson, this methodology presents a subject with a series of triads $\langle i,j,k \rangle$, asking which is the "most similar" and "least similar" within each. Call the number of triads in the series N_t. The crucial premise here, that such pairings are reducible to distance comparisons, is vindicated by data such as those collected by the Project on Occupational Cognition [Coxon *et al*, 1975], where independently-obtained estimates of element similarity produced the same configuration as triadic data. However, at the end of this chapter we will encounter situations in which a minority of subjects make their judgments on avowedly non-geometrical, non-distance-based bases.

The same questions can be phrased less abstractly in terms of "odd-one-out". The subject selects element i as the odd one out of $\langle i,j,k \rangle$, which is equivalent to making the two judgments

$$(i,j) \gg (j,k), \qquad (i,k) \gg (j,k).$$

Call this the 'primary comparison'. The subsequent selection, from the remaining two elements, of k as closer to i (the 'secondary comparison') amounts to making a third judgment

$$(i,j) \gg (i,k).$$

I defer discussion of variant triadic tasks till later in this chapter.

A common use of this triadic task is for mapping colour space [Bechtel, 1976; Helm, 1964; Krantz, 1967; Messick, 1956; Stalmeier & de Weert 1991a, 1991b, 1994; Torgerson, 1956; Wright, 1965]. It has also been applied to:

- visual space [Indow, 1968]
- visual textures [Harvey & Gervais, 1981]
- facial expressions [Gladstones, 1962a, 1962b]

- three-dimensional shapes [Arabie, Kosslyn & Nelson, 1975]
- plant species (varieties of gourd) [Berlin, Breedlove & Raven, 1968].

Not all triad studies have been of elements capable of being presented simultaneously:

- odours [MacRae, Rawcliffe, Howgate & Geelhoed, 1992]
- baby cries [Lyons, Kirkland, Castle & Lawoko, 1991]
- musical intervals [Levelt, van der Geer & Plomp, 1966]
- vowels [Hansen, 1967; Pols, van der Kamp & Plomp, 1969]
- complex tones varying in phase [Hall & Schroeder, 1971; Plomp & Steeneken, 1969]
- musical instruments (timbre space) [Plomp, 1970]
- organ stops [Plomp, 1970]
- complex tones simulating the place-dependent effects of acoustics [Plomp & Steeneken, 1973].

Examples with semantic rather than sensory items include:

- kinship terms [Burton & Nerlove, 1976; Nerlove & Burton, 1972; Romney & D'Andrade, 1964]
- personality traits [Kirk & Burton, 1977; Miller, 1974]
- digits [Miller & Gelman, 1983; Lyons *et al*, 1991]
- animal names [Henley, 1969]
- countries [Dong, 1983; Klingberg, 1941]
- occupations [Coxon *et al*, 1975].

A feature common to all these is the difficulty of quantifying how dissimilar two items are. Each dyad in turn is *sui generis*; informants cannot describe its dissimilarity as a multiple of a mental unit, nor as a fraction of some maximum imaginable dissimilarity. The researcher compensates for the lack of internal units by providing an external basis for a "greater than / less than" comparison, in the form of a second dyad. Three elements suffice for the comparison (the dyads having one element in common): thus the triadic procedure.
It has been used with children [Miller & Gelman, 1983; Seitz, 1971, with 5-year-old subjects]. The odd-one-out phrasing of the notion of triadic combinations comes naturally to young subjects, who baulk at the task of describing dissimilarity in numeric or distance terms (nor should we impose a distance metaphor on them. By assumption, the mental representation of the

stimuli under investigation can be modelled geometrically, so the result of what should be a test of this assumption is prejudiced if it is implicit in the questions). The usefulness to developmental psychologists is obvious.

Another group to benefit are anthropologists, whose attempts to map their informants' mental representations are hindered by linguistic rather than developmental barriers [Berlin, Breedlove & Raven, 1968; Kirk & Burton, 1977; Truex, 1973]. Even animals can provide triadic data. Pigeons have been trained to associate key-pecking with one particular stimulus, and then presented with two stimuli from which they must 'choose' the one that is closer to the target [e.g. Wright & Cumming, 1971].

My first example comes from a study of baby cries [Lyons *et al*, 1991]. Stimuli were 8 tape-recorded cries, selected from a set of 12 for which a configuration was already available, the dissimilarities having been estimated directly [Green, Jones & Gustafson, 1987] and also derived by rating the cries on scales [Gustafson & Green, 1989]. These stimuli (provided by Green and Gustafson) were digitised and stored in a 'sound library' on a Mac computer, which also ran software to present the triads in a random sequence while storing the subject's responses. Eight subjects took part. I will refer to them with the letters 'B' to 'I'.

With 8 stimuli, 56 different combinations of three are possible. 12 triads out of this 56 were presented three times, once at the beginning to provide practice, and again at the end to check for fatigue or satiation effects degrading the data. Finding no such effects, or differences between the practice triads and the subsequent ones, I include all 80 in the analysis. Table 3.1(a) is excerpted from the raw responses from subject B. Table 3.1(b) lists the same responses, now dissected into putative order relationships. Finally, Table 3.1(c) shows these relationships collated to form an input file for the analysis package MTRIAD. We see that (1,2) has been experienced as more dissimilar than (2,3). (1.3) is more dissimilar than (1,2), (2,3), (3,6), and so on.

Table 3.1(a) Excerpt from raw data file
Triad 7 1 5, oddity 2, similarity 1, delay 40
Triad 4 6 1, oddity 3, similarity 2, delay 81
Triad 6 4 3, oddity 2 similarity 1, delay 27
Triad 4 2 8, oddity 1, similarity 3, delay 22
Triad 5 1 4, oddity 2, similarity 1, delay 28
Triad 5 2 1, oddity 1 similarity 2, delay 23
Triad 8 7 2, oddity 2, similarity 1, delay 20
Triad 4 7 8, oddity 1, similarity 2, delay 22
Triad 2 1 4, oddity 3 similarity 1, delay 16
Triad 8 5 6, oddity 1 similarity 2, delay 11
.....

Table 3.1(b) Implied order relationships
(1,7) » (5,7), (1,5) » (5,7), (1,5) » (1,7)
(1,4) » (4,6), (1,6) » (4,6), (1,4) » (1,6)
(3,4) » (3,6), (4,6) » (3,6), (3,4) » (4,6)
(2,4) » (2,8), (4,8) » (2,8), (2,4) » (4,8)
(1,4) » (4,5), (1,5) » (4,5), (1,4) » (1,5)
(1,5) » (1,2), (2,5) » (1,2), (1,5) » (2,5)
(2,7) » (2,8), (7,8) » (2,8), (2,7) » (7,8)
(4,8) » (7,8), (4,7) » (7,8), (4,8) » (7,8)
(1,4) » (1,2), (2,4) » (1,2), (1,4) » (2,4)
(6,8) » (5,6), (5,8) » (5,6), (6,8) » (5,8)
.....

Table 3.1(c) Excerpt from B.LDC, a file of order relationships formatted for MTRIAD

```
8 stimuli
1 2          is greater than...
2 3,

1 3          is greater than...
1 2, 2 3, 3 6,

1 4          is greater than...
1 2,1 3,1 5,1 6,1 8, 2 4, 3 4, 4 5, 4 6, 4 7, 4 8,
```

```
1 5          is greater than...
1 2, 1 3, 1 6, 1 7,1 8,2 5, 3 5,4 5,5 6, 5 7, 5 8,

             .......
6 8          is greater than...
1 8, 3 6, 3 8, 4 6, 5 6,

7 8          is greater than...
1 8, 2 8, 3 8, 6 8,
```

The previous chapter described most details of the MTRIAD algorithm (equivalent to a generalised form of Johnson's "pairwise non-metric MDS"). Recall the Stress, (2.9):

$$S(X) = \sum_{(i,j)\,(k,l)} \varepsilon_{ij,kl}\, H(d_{kl} - d_{ij})\, (d_{kl} - d_{ij})^2 \;\Big/\; \sum_{(i,j)\,(k,l)} \varepsilon_{ij,kl}\, (d_{kl} - d_{ij})^2$$

Following the precedent of Guttman's "soft squeeze" route to minimisation, I treated the denominator as constant, neglecting terms in the gradient coming from its dependence on X. The loss of rigor is more than compensated by the simplicity in the computations: the gradient is now

$$\partial S(X) / \partial x_{ip} = 2 \left(\sum (x_{ip} - x_{jp})/d_{ij} \sum \varepsilon_{ij,kl}\, \Theta(d_{kl} - d_{ij}) \right) \;/\; \text{constant denominator.}$$

Imagining the pairs of elements as linked by "smart springs" clarifies the best order for summations within this formula. MTRIAD iterates over putative order relationships first, comparing each with the reconstructed distances d_{ij} and d_{kl}. Every discrepency, $d_{ij} < d_{kl}$, adds an increment of $2(d_{kl} - d_{ij})$ to the spring force pushing i and j apart, while decrementing the (k,l)-th force by the same amount.

With all contributions to the spring forces summed, the program iterates over springs (dyads), converting each force in turn into contributions to the net forces acting on its endpoints. Each force f_{ij} is exerted parallel to the spring, but it can be resolved into components along the p-th axis which yield, when summed, the components of the net force on i:

$$f_i = \sum_{j=1}^{N} f_{ij} \qquad = \sum_{p=1}^{P} \hat{e}_p \sum_{j=1}^{N} f_{ij} (x_{ip} - x_{jp})/d_{ij} \tag{3.1}$$

where $\hat{e}_p$ is the unit vector parallel to the p-th axis.

So far no constraint has been placed on the overall scale of the configuration. X may change in scale as successive ΔX are added, expanding indefinitely or collapsing to a single point, without altering the Stress. For ease of display it is worth imposing some arbitrary scale. Define "scale", in this context, as the sum of distances squared. Within configuration space there is a surface of configurations having a particular value, C, for this scale; we want X to lie on this surface:

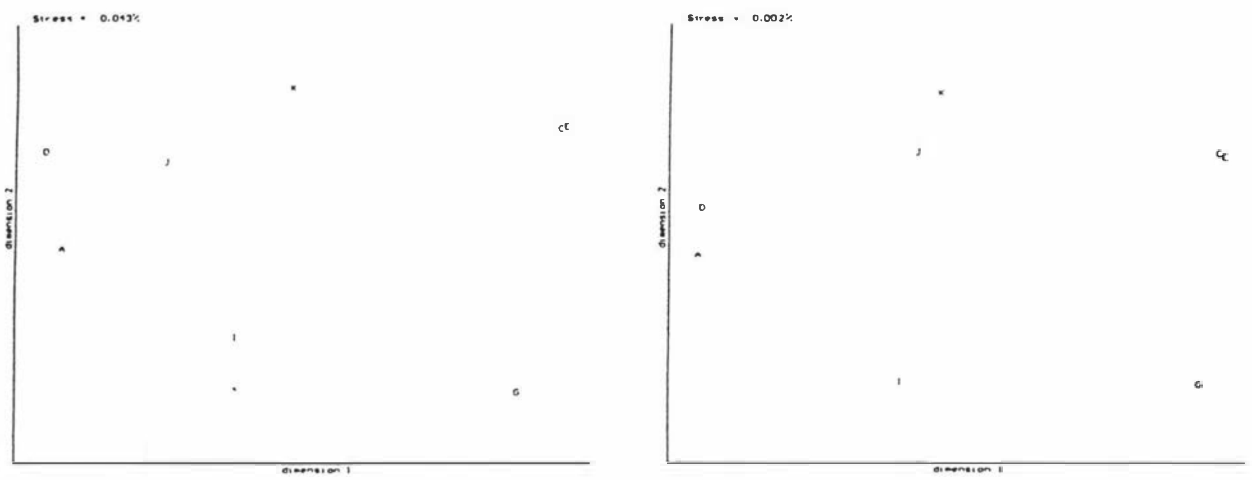
$$\text{scale}(X) = \sum_{(i,j)} d_{ij}^{\,2} = \sum_{p=1}^{P} \sum_{(i,j)} (x_{ip} - x_{jp})^2 = \text{C}. \tag{3.2}$$

$X^{(0)}$ is placed on the surface by multiplying all coordinates by an appropriate scale factor. Each subsequent $\Delta X^{(t)}$ vector can be resolved into a component tangential to the constant-scale surface, and one radial to it; remove the second component.

$$\Delta' X^{(t)} = \Delta X^{(t)} - g\,(\Delta X^{(t)} \cdot g) \qquad \text{where } g \text{ is the unit vector perpendicular to the surface.}$$

Even using $\Delta' X^{(t)}$ instead of $\Delta X^{(t)}$, it proves necessary to periodically renormalise the configuration, since the constant-scale surface is curved: $\Delta' X^{(t)}$ may be tangential, but if its length is non-zero, scale$(X + \Delta' X^{(t)})$ may differ from scale(X).

Figure 3.1 MTRIAD (a, left) and vote-counting (b, right) solutions for subject B



After all this, the MTRIAD solution in two dimensions for the data from B is shown in figure 3.1(a). For comparison, figure 3.1(b) shows the vote-count solution described in chapter 2.

In the MTRIAD solution, the separation into three distinct types of cry, (A, D, J; C, E, K; and G, I), as shown in an independent MDS analysis for the same stimuli [Green, Jones & Gustafson, 1987], is clearer than in the vote-count solution.

I note for later reference that this data base also contains confidence ratings provided by subjects H and I. A refinement in the computer interface allowed them to rate their degree of confidence, after every primary and secondary comparison they made, as 1, 2 or 3. High confidence ratings are associated strongly with large differences between the MTRIAD reconstructions of the distances being compared. This is shown, for subject H's data, in figure 3.2. The distance difference is $|d_{ij} - d_{ik}|$, for secondary comparisons, while in the case of primary comparisons, the distance difference plotted against the confidence is the average of $|d_{ij} - d_{jk}|$ and $|d_{ik} - d_{jk}|$.

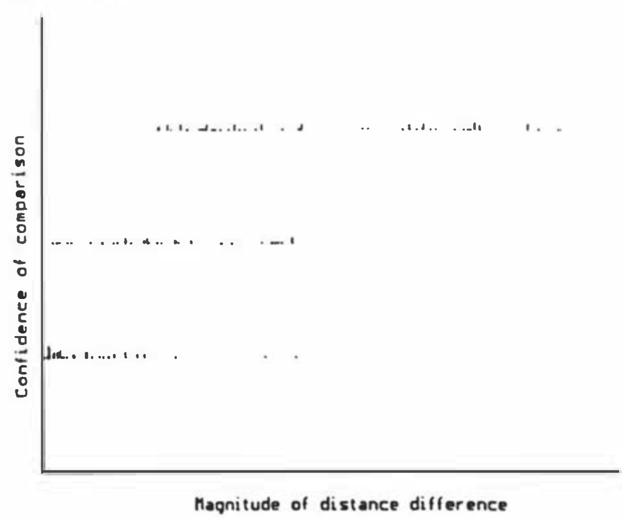


Figure 3.2 Distance difference against confidence

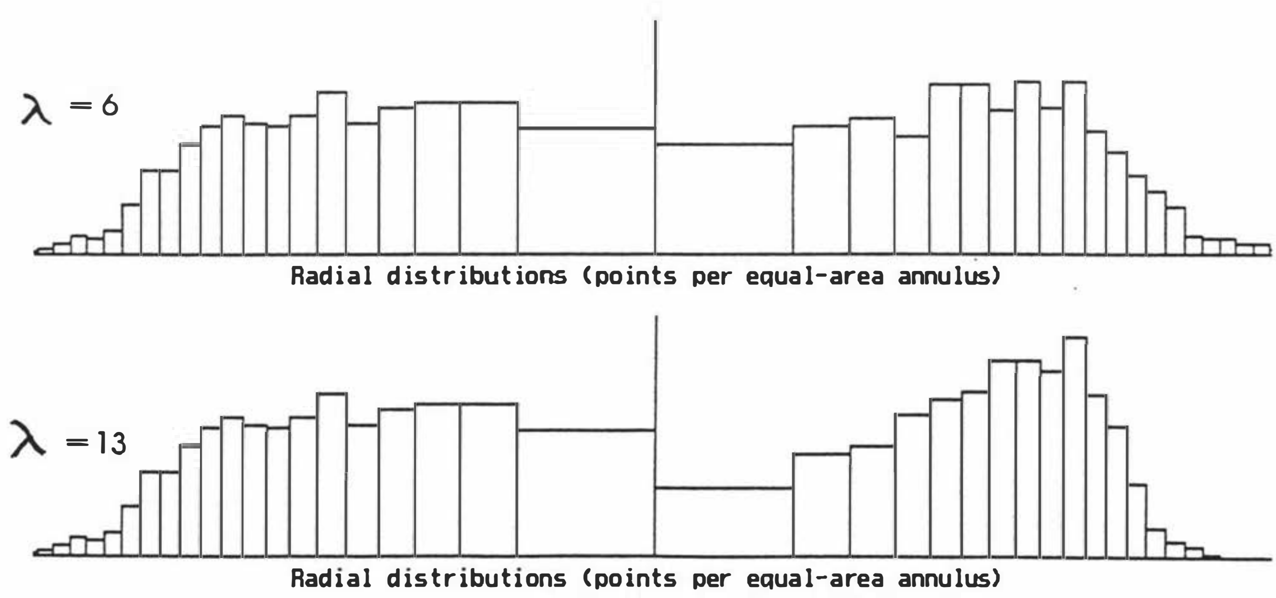
Chapter 2 discussed some theoretical problems with the vote-count procedure. In practice, it may be that the issue is academic, with any distortions it adds to the rank order disappearing again during the scaling process. We find that inserting vc_{ij} into (2.9) produces a value of 0.0026 for

Stress (these low Stresses reflect the small, marginal numbers of stimuli and triads). Embedding vc_{ij} in a two-dimensional space produces d^*_{ij}, the distances of figure 3.1(b), which yield 0.0075 when inserted in (2.9). Finally, the Stress for the d_{ij} of the MTRIAD solution is 0.0004.

It seems that the more elements, and the more evenly they are distributed in the configuration, the better the vote-count approximation (here I am generalising from observations, rather than a systematic exploration).

Systematic distortions might not show up in correlation statistics. I conducted a Monte Carlo experiment, generating synthetic triadic data for 100 random configurations and reconstructing them with the vote-count procedure. N was 15. Radial coordinates of elements are accumulated and plotted, for the actual and reconstructed configurations, in the left and right panels respectively of figure 3.3. There are signs of a "centrifugal effect": points tend to migrate toward the periphery. This agrees with Gladstones' obervation of "...the tendency of the approximate methods [for analysing triadic data] to overestimate distances between stimuli near the centroid." [1962b, page 200], and also with observations of the distortions produced by low-resolution dissimilarity estimates [Green & Rao, 1971; also Borg & Lingoes, 1987]. Viewed with a sufficiently unskeptical eye, figure 3.1(b) reveals the same distortion.

Figure 3.3 Tendencies for elements to appear in particular radial coordinate bands, for 100 randomised configurations (left) and their vote-counted reconstructions (right)



Another approach to analysing triadic data, suitable when the triads have been judged repeatedly (not necessarily by the same judge), comes to us from Torgerson [Gladstones, 1962a, 1962b; Torgerson, 1952, 1956]. Torgerson's name is also linked with a variant triad procedure, but I prefer to describe his method of processing the data separately from its collection. The method combines variability in the replications with arguments borrowed from one-dimensional scaling theory, to achieve interval-level data, grist for metric MDS. Briefly, if two dyads (i,j) and (j,k) have been repeatedly compared, Torgerson invokes the Law of Comparative Judgment and applies a probit transform (the inverse of the cumulative density function). The result is an interval relationship between the two distances: $d_{ij} = d_{jk} + a(p)$, where p is the proportion of the comparisons in which $(i,j) \ll (j,k)$ and the function $a(p)$ is known.[1]

For i, j and k, points defining a triangle in a mental map which we wish to reconstruct, responses for a single triad are a statement about the triangle's orientation: which corner is acutest and which most obtuse. With this approach we learn more about its geometry.

These interval relationships form a set of simultaneous equations. The second phase of Torgerson's analysis performs a least-squares calculation to produce an approximate, low-dimensional solution for the coordinates x_{ip}.

As a result, triadic data could be scaled before the advent of non-metric tools rendered proximity data equally tractable. This, as much as the advantages described earlier, accounts for the popularity of the triadic method during the 1950s and 1960s. Torgerson's method is of more than historical interest, and is summarised here to foreshadow the use of similar arguments in Chapter 5.

After that digression, one commonly-used alternative analysis remains to be described; MINITRI [Roskam, 1970], also implemented as TRISOSCAL, part of the Cambridge package for MDS [Coxon & Jones, 1978; MacRae, Howgate & Geelhoed, 1990].

MINITRI is a version of the KYST algorithm, tailored to handle triadic data: the averaging of distances which violate the rank-order of the data, yielding disparities, is performed only within

[1] Compare with Klingberg's use of the logit function, to transform vote-count estimates (summed over many respondents) into a form suitable for metric analysis (principal coordinates).

triads. Reconstructed distances are simultaneously adjusted to improve their fit to a separate list of disparities for each triad in the data set. It is instructive to inspect the stress contributions from each triad in isolation. Let the l-th triad be <i,j,k>, with its elements ordered so that

$$(i,k) \gg (i,j) \gg (j,k).$$

The corresponding l-th list of disparities is $\eth_{ik:l}$, $\eth_{ij:l}$, $\eth_{jk:l}$.

When the rank order of reconstructed distances is a complete reversal of the data, $d_{ik} < d_{ij} < d_{jk}$, then

$$\eth_{ik:l} = \eth_{ij:l} = \eth_{jk:l} = (d_{ik} + d_{ij} + d_{jk}) / 3,$$

and the contribution to raw stress from triad l is then

$$(d_{ik} - \eth_{ik:l})^2 + (d_{ij} - \eth_{ij:l})^2 + (d_{jk} - \eth_{jk:l})^2$$

$$= \{d_{ik} - (d_{ik} + d_{ij} + d_{jk}) / 3\}^2 + \{d_{ij} - (d_{ik} + d_{ij} + d_{jk}) / 3\}^2 + \{d_{jk} - (d_{ik} + d_{ij} + d_{jk}) / 3\}^2$$
$$= \{(d_{ik} - d_{ij})/3 + (d_{ik} - d_{jk})/3\}^2 + \{(d_{ij} - d_{ik})/3 + (d_{ij} - d_{jk})/3\}^2 + \{(d_{jk} - d_{ik}) /3 + (d_{jk} - d_{ij})/3\}^2$$

$$= 2/9\,(d_{ik} - d_{ij})^2 + 2/9\,(d_{ik} - d_{jk})^2 + 2/9\,(d_{ij} - d_{jk})^2$$
$$+ 2/9\,(d_{ik} - d_{ij})\,(d_{ik} - d_{jk}) + 2/9\,(d_{ik} - d_{ij})\,(d_{ij} - d_{jk}) + 2/9\,(d_{ik} - d_{jk})\,(d_{ij} - d_{jk})$$

$$= 1/3\,(d_{ik} - d_{ij})^2 + 1/3\,(d_{ik} - d_{jk})^2 + 1/3\,(d_{ij} - d_{jk})^2. \qquad (3.3)$$

(3.3) is equal (apart from the factor 1/3) to the contribution of such a triad to the MTRIAD stress.

If the reconstructed configuration clashes with just one of the inequalities – the first, for example – then:

$$\eth_{ik:l} = \eth_{ij:l} = (d_{ik} + d_{ij}) / 2; \qquad \eth_{jk:l} = d_{jk}.$$

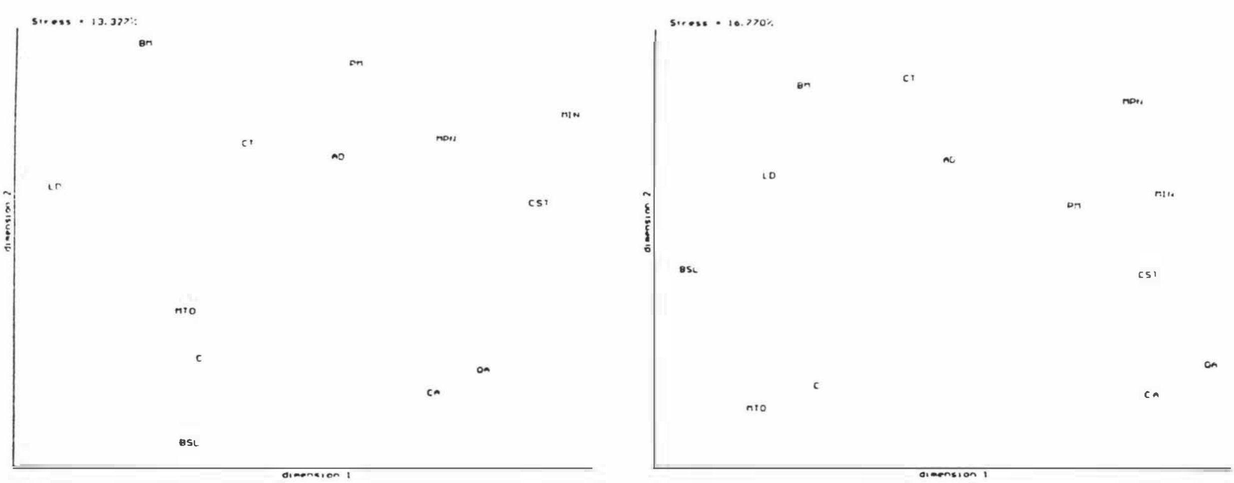
and the contribution to raw stress becomes

$$(d_{ik} - \eth_{ik:l})^2 + (d_{ij} - \eth_{ij:l})^2$$
$$= ((d_{ik} - d_{ij}) / 2)^2 + ((d_{ij} - d_{ik}) / 2)^2 = (d_{ik} - d_{ij})^2 / 2. \qquad (3.4)$$

Again, (3.4) boils down to the same contribution to Stress (and therefore to restorative forces), albeit with a different scale factor.

For confirmation of the closeness of the two programs, I applied MTRIAD to data from a triadic study of odours (kindly provided by MacRae), yielding a solution which agreed with that published by MacRae *et al.* [1992]. A second source of confirmation is the POOC data, which includes triads from 47 subjects for 13 names of occupations. MTRIAD duplicates the Coxon and Jones solution [1978, p. 77]. In this case, the responses varied from subject to subject, and often were fraught with internal contradictions within one subject's data, leading to higher Stresses. The Stress from inserting vc_{ij} into (2.9) is 0.1116. This becomes 0.1498 for d^*_{ij}, the result of scaling the vc_{ij}, i.e. embedding them in two dimensions (figure 3.4(b)). Finally, the Stress in the MTRIAD solution, figure 3.4(a), is 0.1333.

figure 3.4 MTRIAD (a, left) and vote-count (b, right) solutions for occupational-title data (Coxon *et al*)



Key to the occupations:		AD	Ambulance driver
BM	Barman	BSL	Building-site labourer
C	Carpenter	CA	Chartered accountant
CT	Commercial traveller	CST	Comprehensive school-teacher
LD	Lorry driver	MIN	Church of Scotland Minister
MPN	Male Psychiatric Nurse	MTO	Machine tool operator
PM	Policeman	QA	Qualified actuary

A possible advantage of the MINITRI approach over MTRIAD is that the former can be extended to Guttman's rank-image disparities. In other respects it is less flexible. Forms of data other than the Method of Triadic Comparisons require different programs from the MINISSA series.

Variations on the Triadic Theme

Andrews and Ray recommend making the triad procedure less onerous by omitting the secondary comparisons. "It appears that adequate information may be gathered from a triad by a single though more complex judgment if the S is requested merely to select from each triad the stimulus that does not 'belong'." [Andrews & Ray, 1957]. Theirs was a factor-analytic context, but for an MDS confirmation of the simplified tasks's validity, see Gladstones [1962b] – a vote-count analysis. Upon first glance, it might seem that the varying number of comparisons involving different dyads would be a source of distortion when vote-counting them, but the effect is systematic, and non-metric methods correct it.

Anthropologists are common users of this variation [Berlin, Breedlove & Raven, 1968; Borgatti, 1991; Burton & Nerlove, 1976; Kirk & Burton, 1977; Nerlove & Burton, 1972; Romney & D'Andrade, 1964; Truex, 1973], though they phrase the question differently, asking for the "most similar pair", which implies the same two comparisons.

Coxon and Jones are dismissive of it. Nevertheless, the procedure has advantages: for only half as many responses, two-thirds of the comparisons are provided. Moreover, those choices are often disproportionally easy (though not always so. One can imagine three stimuli positioned at roughly equal intervals along a straight line in the mental map, so that once the odd-one-out is chosen, the central stimulus stands out as more similar to it, but making that initial choice is difficult). There is indirect evidence for asserting that the odd-one-out judgment is easy to make, in the form of the confidence ratings from subjects H and I in the Lyons *et al* cry data. The primary comparisons were made with confidence, and vice versa (the data are not complete for subject I, who was called away before finishing all the triads).

	Subject H			Subject I		
Confidence	1	2	2	1	2	3
Primary comparisons	3	9	68	4	23	33
Secondary comparisons	51	28	1	31	29	0

The effects of omitting secondary comparisons from the cry and occupation data are minor. But some caution is required. If elements happen to be arranged in an elongated strip, so that the major dimension(s) dominate the primary comparisons, then secondary comparisons are vital for resolving separations along the minor axis, which can disappear without them.

The focus so far has been on the Method of Triadic Comparisons. In the "Complete method of triads" [Torgerson, 1952], the three comparisons per triad are made separately. The subject is presented with the triads, usually in a randomised order, and asked "Is i or k closer to j?" for each $<i,j,k>$: the other two comparisons ((i,j) against (i,k), and (i,k) against (i,k)) being separate questions. Examples of this procedure are Arabie, Kosslyn and Nelson [1975], Seitz [1971], Stalmeier and de Weert [1991b, 1994]. The nomenclature is unclear, with some authors reserving the name "complete method of triads" for Torgerson's algorithm, described earlier. Coombs [1964] objects to both phrases, "method of triadic comparisons" and "complete method of triads", and puts forward "Method of similarities" for the former, regarding it as a specific case of the more general 'cartwheel' method.

The same format that was used with the Method of Triadic Comparisons data is used again to collate data obtained this way into input files for MTRIAD. However, the methods are not completely equivalent. In the former one, there are $2^3 = 8$ possible ways of responding, compared to only 6 in the latter method, where the primary comparison restricts possible secondary comparisons. Analyses such as Torgerson's, which involve Thurstone's Law of Comparative Judgment axioms, and rely on the comparisons being independent, require the Complete Method of Triads for full rigor.

Stalmeier and de Weert [1991a], describe an ingenious variant, relying on gestalt fusion, which enables subjects to cope with $N = 16$ stimuli (colours) and $N_t = 560$, i.e. 1680 separate comparisons. For each comparison, stimulus j is presented on a computer display, as a solid hexagon of colour, flanked by six triangles (three each, alternating, of colours i and k), creating a star-of-David pattern. The eye simplifies this pattern by fusing the hexagon with whichever set of triangles are closer in colour, to form a larger triangle: the other three triangles being excluded. The comparison is phrased as the question, "which way is the large triangle pointing", which can be answered without conscious effort (figure 3.5).

Presumably this would work as well with visual textures [Harvey & Gervais, 1981].

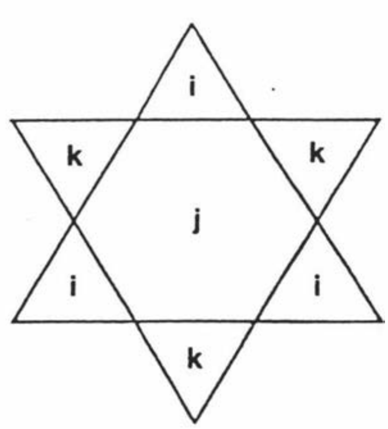
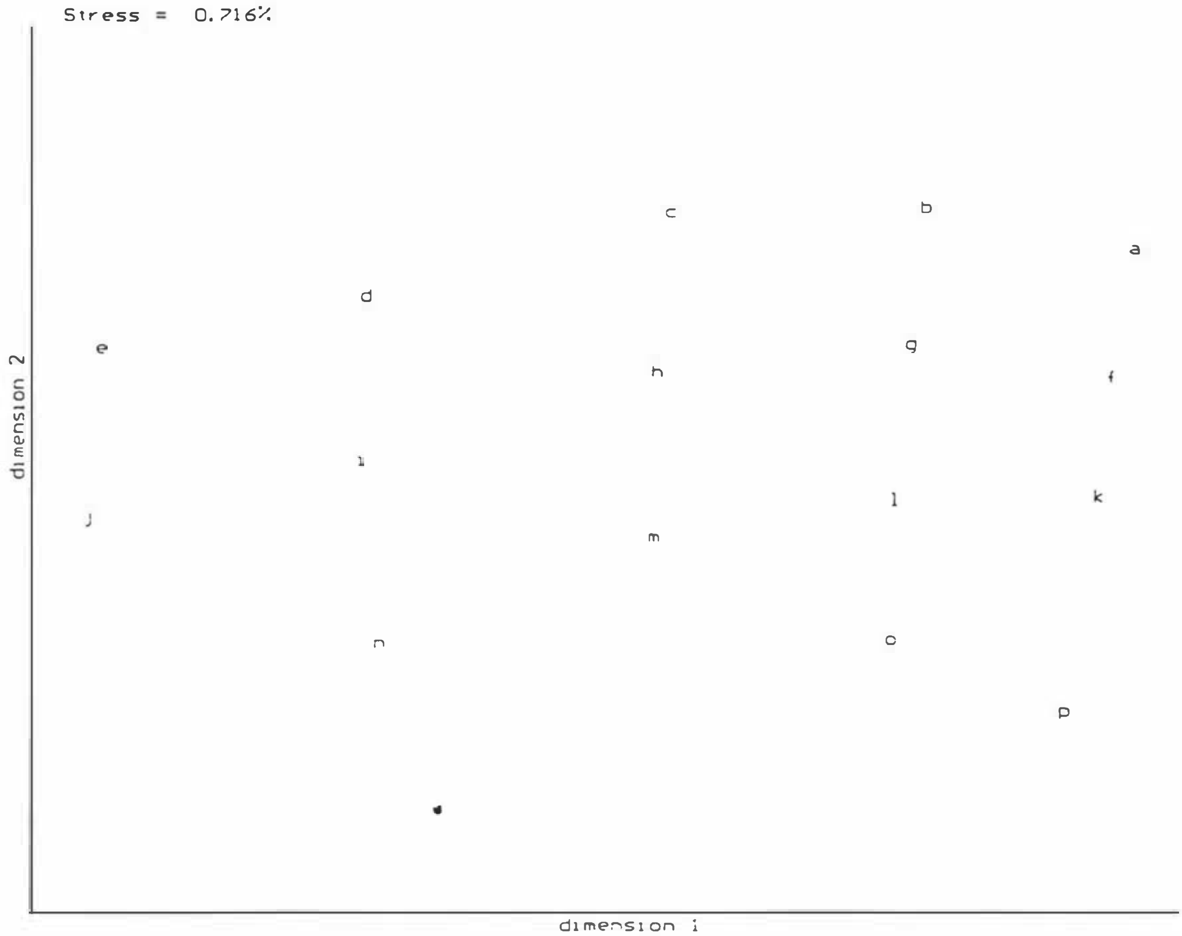


Figure 3.5 The 'Star-of-David' design

Figure 3.6 Two-dimensional solution to Stalmeier-de Weert triadic data for subject PE



Stalmeier generously provided me with copies of his data. Figure 3.6 is an MTRIAD analysis of the data for subject PE. The differences from the vote-count solution are negligible. The positions of these stimuli within colour space are shown in Stalmeier and de Weert [1991a]. Luminance was constant, so there are two dimensions (reflecting the conversion, in the retina, of cone-cell signals into a pair of opponent-process signals). There is a red-green axis, intersected by tritanopic confusion lines running roughly in a blue-yellow direction. Compared to the configuration which Stalmeier and de Weert obtained using more conventional ways of presenting triads [1991b], the red-green axis is more pronounced. Perhaps colour dissimilarity, as measured by the tendency of adjacent areas of colour to fuse, conflates two things. Edge distinctness is not identical with perceived colour dissimilarity [Tansley & Valberg, 1979]. It is known that the *parvo* cell pathway which mediates edge discrimination responds poorly to blue-yellow differences, compared to the *magno* pathway mediating colour recognition [Livingstone, 1988]. Anyone planning to replicate the Stalmeier-de Weert gestalt-fusion work should consider separating the coloured polygons with gaps of background colour, to avoid this problem.

In experiments with this method, I encountered another problem, which might be termed a transparency effect. Besides the two gestalts described, the left- and right-pointing triangles, some combinations of stimuli lead to the perception of a third gestalt, in which one coloured triangle is visible behind a second, partially transparent one, with the hexagon being their area of overlap.

Torgerson suggested modifying the triadic procedure to collect distance ratios directly, to be solved as simultaneous equations. Several researchers took up this proposal [Helm, 1964; Indow, 1968; Wright, 1965], but the variant has since fallen into neglect. The need for it was obviated by Shepard's well-known demonstration [Shepard, 1962] that for a reasonable number of elements ($N > 7$), distance comparisons provide sufficiently stringent constraints on the configuration to position each element with confidence. Once the orientation of each triangle in a map is identified by labelling one corner as the most acute and another as most obtuse, very little added precision comes from knowing its exact proportions.

No-one holds a copyright on the word 'triad', and some of its meanings fall outside the domain of this dissertation.

It straddles the borderline when used in connection with repertory grids. Kelly [1955] pioneered the argument that reliable judgments of separation or quantity require at least three items. Thus triads are used with repertory grids as a way of eliciting scales for rating purposes: the two most similar elements (the construct) forming one pole of the scale, and the odd-one-out of the triad (the "contrast") forming the other. From there, their rôle varies. See, for instance, Tyska and Goszczynska [1993], Vlek and Stallen [1981], in which subjects were asked to sort the elements into two piles, placing each element with either the construct i or the constrast (j,m) (repeating this for eight randomly-picked triads $<i,j,k>$). Although presented as a sorting task, this procedure is closer to a triadic one, not so much in its treatment of $<i,j,k>$, but in the judgment the subject makes for each element $m \notin <i,j,k>$ – whether δ_{im} is greater or less than m's distance from the (j,k) nucleus. Equivalently, it is a low-resolution scale. In some other applications, the scales have higher resolution but are still anchored by the triads which elicited them; these could be regarded as variants of ratio triads.

Bechtel reminds us that *any* scale data can be converted to triadic form, while Coombs makes the same point about preference data: the elements in effect are ranked by their proximity to an implied third point (the scale endpoint, which may be at infinity, in the first case; the "ideal point" in the case of preferences). The data become explicitly triadic if pairwise comparisons are used to order the stimuli along the scale [Ramsay, 1980] or preference gradient [de Soete & Winsberg, 1993]. Processing them becomes a matter of adding additional 'virtual' elements for the end points or ideal points. If stimuli are ranked, it is more a situation of conditional rank ordering, to be discussed in Chapter 6.

Element numbers

What is the optimal number of stimuli? $N = 6$ is too few to reliably recover the configuration: there are only 20 triads, 60 comparisons, not enough constraints on stimulus positions. Bissett and Schneider [1992] recommend 20 $N R$ comparisons for credible recovery. Combined with vote-count artifacts, this shortfall could explain why Hall and Schroeder's study [1971] of six complex tones (differing only in the phase of the two constituent tones) gave results divergent from studies with eight and 15 stimuli [Plomp & Steeneken, 1969] and a small-scale project, shortly to be described, with the author as the subject, where $N = 12$. Seitz used only four items in her research with children [Seitz, 1971] but her interest was in individual differences, not the configuration.

Eight, in my experience, is minimal. Even there, there are only $56 \cdot 3$ comparisons, whereas Bissett and Schneider argue for twice that number, but their criterion is overcautious, if the consistency amongst individual-subject configurations for the data of Lyons *et al* is any guide.

With 10 or more elements, a new problem impinges: the number of triads goes up as the third power of N (specifically, $N_t = N(N\text{-}1)(N\text{-}2) / 6$). Limits to the endurance and good humour of one's subjects impose a ceiling on N.

Dong [1983] found evidence that satiation and fatigue were increasing the fallibility of responses as the 84 triads of a nine-element set of country names progressed. On the other hand, Stalmeier and de Weert's subjects [1991a] viewed each of the 1680 comparisons four times (to use both orders, $<i,j,k>$ as well as $<k,j,i>$, doubled again to check reliability), over the course of three hours, without obvious differences between the first time and the fourth. It may be that conceptual stimuli are more tiring.

In a truly heroic study, Harvey and Gervais [1981] used 30 stimuli (visual textures). However, the 4060 triads were administered in daily sessions spaced over a month, and the two subjects (the authors) were highly motivated.

The triadic method has fallen into relative neglect, as researchers weigh the greater precision it offers against the prospect of large areas of *terra incognita* in their maps. Whatever the triad threshold where open rebellion breaks out in the laboratory, we will always want to scale more stimuli, more landmarks. This raises an issue: can the list of triads be thinned out, somehow?

There are encouraging precedents, in the experimental designs for dissimilarity data which allow a larger N by treating some proportion of the dyads as missing data [Spence & Domoney, 1974]. A finer-grained approach is possible here: instead of neglecting a dyad, one omits a proportion of its comparisons with other dyads. Chapter 8 looks in detail at the choice of *which* to omit.

Balanced Incomplete Block Designs (BIBDs) are widely used [e.g. Arabie, Kosslyn & Nelson, 1975; Burton & Nerlove, 1976; Kirk & Burton, 1977; MacRae, Howgate & Geelhoed, 1990]. These involve selecting a subset of the triad list to ensure that each dyad takes part in an equal number of triads, λ (where $\lambda \leq N$-2). A BIBD ($N = 13$, $\lambda = 2$, $N_t = 52$) was used in collecting the POOC data which furnished figure 3.4.

Triadic Experiments with sounds (1)

The four following applications of the Method of Triadic Combinations all used synthesised sounds, reproduced on a Voicecard 1000 sound board, manufactured by Teleste Oy, installed in a 286 PC. The Voicecard 1000 samples sounds at 20 kHz and 8-bit precision. The board and its associated software driver were intended for digitising and storing sound input, for playing back later, but a little trial-and-error revealed the format used for the storage. This enabled me to write simple additive-synthesis programs to create files which contained the waveforms of novel sounds, including ones which would be hard to produce in any other way.

The board is monophonic. In three experiments, its output was connected to a stereo amplifier, with the 'mono' switch set so that the subject could adjust the balance of the volume from the two speakers to a comfortable level. These experiments were conducted at home and created consternation among the neighbours when they found themselves apparently living next door to an avant-garde musician. In the third experiment to be described (chronologically, it was the first), amplification of the board's output was with a tape-deck, which lacked a 'mono' switch, and subjects heard the sounds through one or other earpiece of a headphone set: it was not possible to equalise the output through both earpieces at once. This is a source of variation (neglected in the analysis), since depending on their hemispheric laterality, people differ in which ear is most sensitive to the details of complex sounds. Also ignored are diurnal variations in auditory acuity.

The 12 sounds used in the first experiment were synthesised using tones which were carefully chosen to remove the sense of a particular pitch from the combinations, while leaving each one with a distinct musical key ("chroma"). These ingenious sounds were first described by Shepard [1964] so in what follows I call them "Shepard tones". They are comprised of power-of-two harmonics and sub-harmonics of the frequncy of the key. If that frequency is f_1, then the components are $f_h = f_1 \cdot 2^h$, where $h \in \{ \dots -3, -2, -1, 0, 1, 2, 3, \dots \}$. The list of harmonics and sub-harmonics is prevented from extending indefinitely into ultrasonic and infrasonic frequencies by a cosine amplitude envelope, peaking at f_p, with γ being the number of octaves spanned:

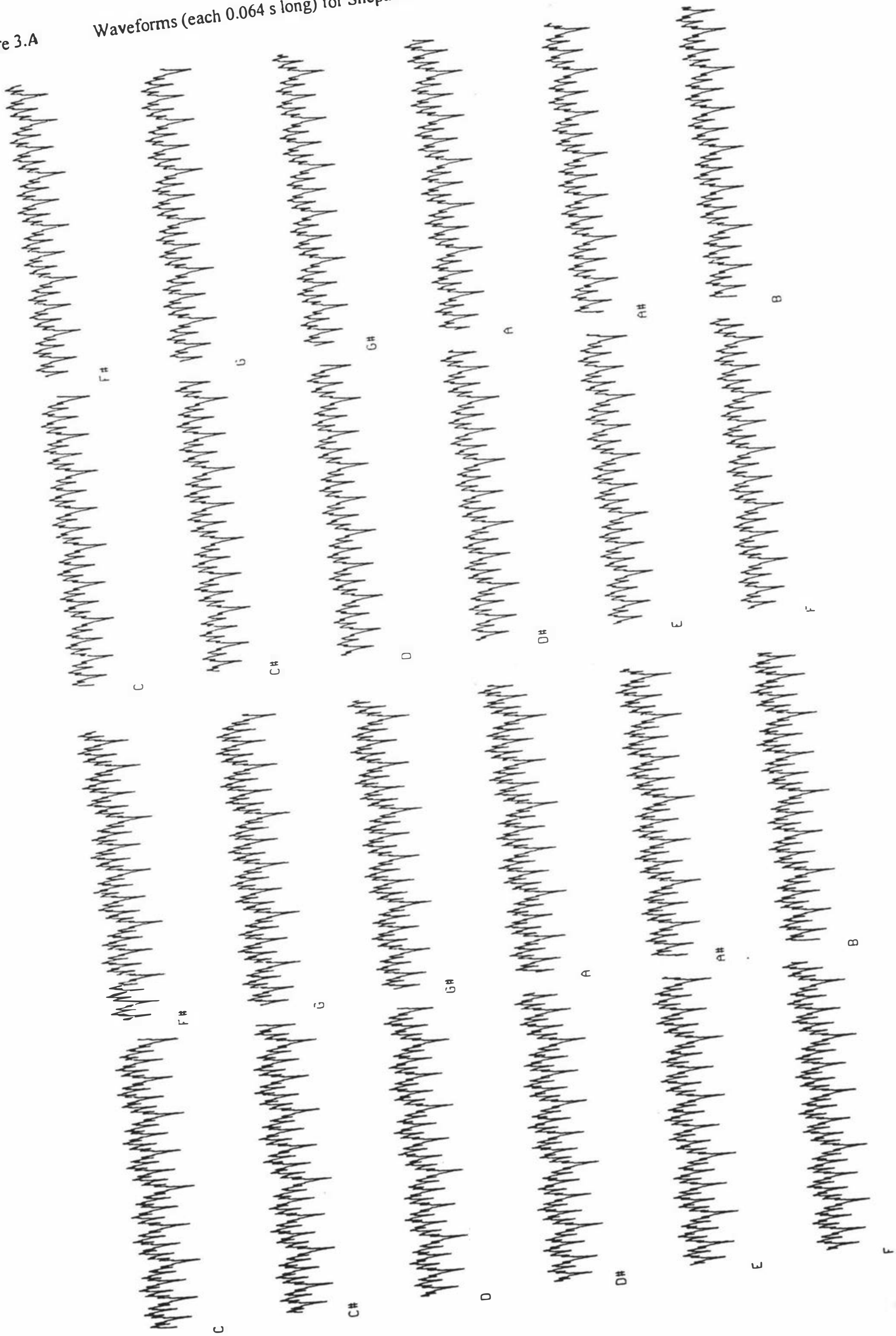
$$a_h = a_1 \{1 + \cos(2\pi \log_2(f_h / f_p) / \gamma)\} / 2 \qquad (3.5)$$

$$\text{for } f_{min} \le f_h \le f_{max}, \text{ where } f_{min} = f_p\, 2^{-\gamma/2},\ f_{max} = f_p\, 2^{\gamma/2}$$

Then the combined waveform is $A(t) = \sum_h a_h \cos(2\pi\, f_h\, t)$.

The resultant tones are notes as opposed to chords. Each has a rich, organ-like quality, undefinable in octave, comparable to simultaneously striking all the C keys (for instance) on a piano keyboard.

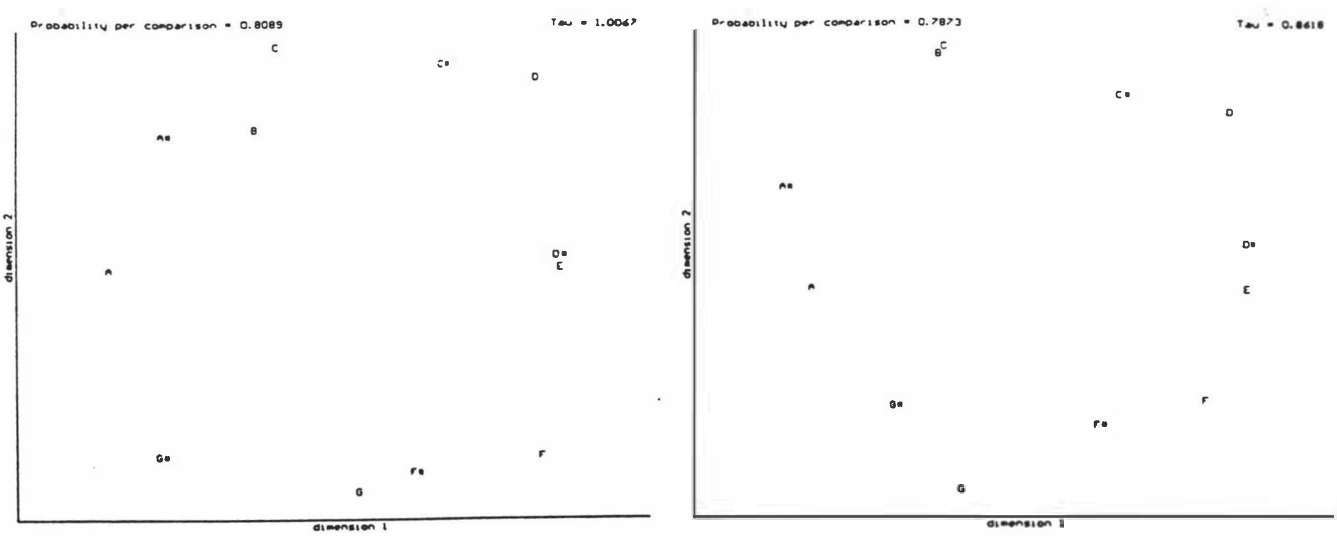
For the 12 tones synthesised for this experiment, the peak frequency f_p was 400 Hz, and f_1 took on the frequencies of the 12 semitones within the middle octave (well-tempered scale). Each gure 3.A Waveforms (each 0.064 s long) for Shepard tones with $\gamma = 8$ (top) and 10 (bottom).



stimulus lasted for 1 second, with a_l constant for the central 0.8 s, tapering to zero during the first and last 0.1 s to prevent transients. The subject (the author) adjusted the stereo amplification to a comfortable level. Details of the computer interface collecting responses to each triad are given in Chapter 8; for now, the salient points are that each triad was presented by playing the three stimuli, in a randomised order, separated by 0.5 s pauses; while judging them the subject could replay any of the stimuli as often as needed. The 220 triads for $N = 12$ were reduced to 132 with a Balanced Incomplete Design ($\lambda = 6$).* I underwent this experiment twice, once providing both primary and secondary comparisons, and the second time providing primary comparisons only. The replications also differed in the number of octaves spanned by the stimulus sets, with $\gamma = 8$ in the first replication (i.e. $f_{min} = 25$ Hz, $f_{max} = 6.4$ kHz) and $\gamma = 10$ in the second ($f_{min} = 12.5$ Hz, $f_{max} = 12.8$ kHz). Segments (0.064 s long) of the resulting waveforms are plotted in figure 3.A.

Although there were subtle timbre differences between the tones, I found that my main criterion for dissimilarity was "perceived pitch difference". Hearing two stimuli close in key, e.g. C and D, the illusion of a *pitch* difference is very strong. In fact, average pitch is the same for all stimuli, and the perceived pitch ranking is not transitive: there are tones X, Y, Z such

Figure 3.B Configurations for 12 Shepard tones with $\gamma = 8$ (left) and 10 (right)



---

* For Balanced Incomplete Designs, see pages 182-185; also page 38.

that X sounds higher than Y, Y higher than Z, and Z higher than X.[2] They are no more "high" or "low" than 11 on a clock face is earlier or later than 1 (of what day?).

The triads boil down to the panels of figure 3.B. Both exhibit the expected circular sequence of the tones. Shepard estimated the dissimilarity between pairs of tones indirectly (for $N = 10$) and obtained a similar circular arrangement. Basically the configuration is one-dimensional, but it requires a second dimension in order to loop back on itself, like hours on a clock-face. In a serious investigation of the properties of these stimuli, it would be possible (and desirable) to incorporate the cyclic nature of the "pitch" dimension into the definition of "distance", thus eliminating the need for the second dimension. As it is, if the stimuli *do* vary significantly along some second dimension, accommodating such differences in the configuration would involve displacing points away from the circle along a *radial* dimension (with non-Euclidean effects on the distances between them), or else would require a third dimension.

Perhaps the departure from circularity in figure 3.B reflects some second dimension (though it seems equally plausible to ascribe it to the incompleteness of the data).

A final remark about figure 3.B: it was produced using a *Maximum Likelihood* generalisation of the stress-minimising algorithm, the subject of Chapter 5. This applies also to the remaining experiments to be reported in this chapter.

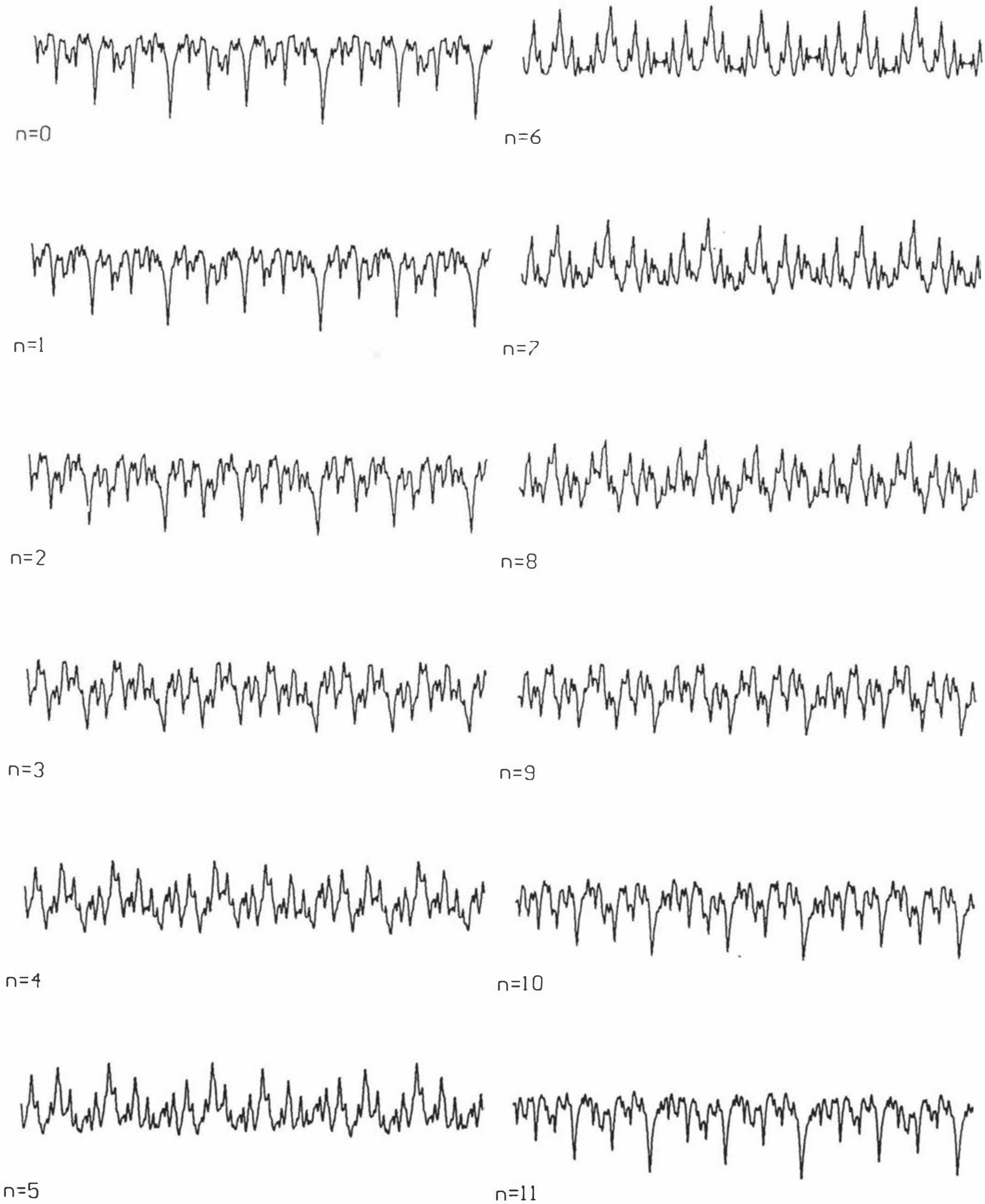
Experiments with Sound (2)

It is known that changing the relative phases between components of a complex tones can affect the tone's timbre to a small, but detectable, extent. Plomp and Steeneken [1969] demonstrated this in an elegant application of the Method of Triadic Combinations. In a series of six experiments, they used eight, nine, and 15 stimuli, each analysable into the same components: the first 10 harmonics of a fundamental frequency (f_1 = 292.4 Hz, except in their experiment 4). In most of their experiments, the amplitudes of successively higher harmonics dropped off according to a slope of -6 dB / octave, i,e, the amplitude of the *h*-th harmonic was $a_h = a_1 / h$. These harmonics produce a triangular or saw-tooth wave-form when combined in

---

[2] Deutsch [1992a, 1992b] and Shepard [1964] have exploited this non-transitivity and devised ingenious auditory illusions: the counterparts of visual illusions.

Figure 3.7 Shepard-tone intervals with 12 different phase shifts between the components (each sample 0.064 s long)



sine phase. To create the other stimuli, Plomp *et al* shifted the phase of various harmonics. They found that all-sine and all-cosine stimuli were very similar,

$$A(t) = \sum_{h=1}^{10} a_h \sin(2\pi\, f_h\, t) \qquad \text{and} \qquad A(t) = \sum_{h=1}^{10} a_h \cos(2\pi\, f_h\, t),$$

as were stimuli composed of alternating sine and cosine terms,

$$A(t) = \sum_{\text{even } h} a_h \sin(2\pi\, f_h\, t) + \sum_{\text{odd } h} a_h \cos(2\pi\, f_h\, t),$$

and $$A(t) = \sum_{\text{even } h} a_h \cos(2\pi\, f_h\, t) + \sum_{\text{odd } h} a_h \sin(2\pi\, f_h\, t).$$

The maximum dissimilarity was between all-sine and all-cosine stimuli on one hand, and alternating-term stimuli on the other, lying at the extremes of the major axis of 'phase space'. They did not attempt to interpret higher dimensions.

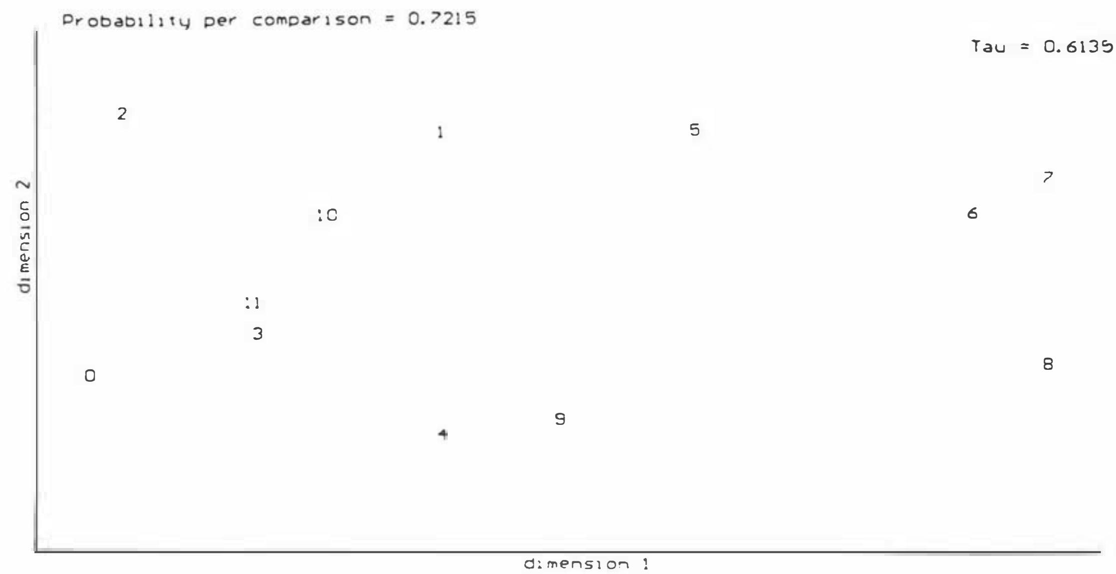
I set out to replicate this work with the same hardware as in the previous experiment. The stimuli were 12 phase-shifted versions of an interval made by combining two Shepard tones ($f_p = 400$ Hz), with $f_1^{(l)} = 400$ Hz and $f_1^{(h)} = 600$ Hz. To generate a simple family of waveforms, parameterised by a single cyclic variable ϕ, I shifted the phase of the second set of harmonics:

$$n\text{-th stimulus} = A(n, t) = \sum_h a_h^{(l)} \cos(2\pi\, f_h^{(l)}\, t) + \sum_h a_h^{(h)} \cos(2\pi\, f_h^{(h)}\, t + \phi_n)$$

where the n-th phase shift $\phi_n = 2\pi\, n \,/\, 12$, for $0 \le n < 12$, and the amplitudes a_h of the components were defined as before by (3.5). Figure 3.7 illustrates the resulting waveforms. The experiment used a $\lambda = 5$ BID (110 triads), with both primary and secondary comparisons. There was only one subject, the author. Subjectively, the differences between the tones were ones of pitch as much as timbre, as if the perceived amplitudes of the various harmonics varied with the phase relationships between them, some 'masking' others, affecting their contributions to the average pitch.

The resulting configuration, figure 3.8, supports Plomp and Steeneken, in that a single dimension dominates the perceived dissimilarities, with the all-cosine stimulus ($n = 0$) at one extreme and $n = 6$ (alternating terms) at the other. This contradicts a triadic study [Hall & Schroeder, 1972], which used only six tones (each consisting of two components), and found two dimensions, with equal salience. A second dimension is shown in figure 3.8; it could be a genuine feature of the phase shifts, but could also be an artifact produced by hardware limitations, or by the incompleteness of the data. Very few conclusions can be drawn from a single-subject study.

Figure 3.8 Configuration for 12 stimuli varying in the phase between components



Intervals – Background

The third and fourth experiments look at simultaneous musical intervals, i.e. tones made up of simpler tones with small-integer ratios between their fundamental frequencies. Simultaneous intervals are also called harmonic; they are distinguished from sequential, or melodic, intervals. Examples are tone combinations with frequency rations of 1:2 and 2:3. These are the Octave and the Third respectively (in "just intonation").

Such ratios are more consonant or euphonious to the average Western ear than larger-integer ratios such as 11:12 or 8:15. According to the prevailing theory, the net dissonance of an interval is a sum of contributions from each of the harmonics of one tone interacting with each of the harmonics of the other. Each harmonic lies in the centre of a range of frequencies, the *critical band*, such that other harmonics in the band interfere to create beats which are perceived as 'roughness' and contribute to dissonance. When the lower of the two fundamentals is a simple fraction of the higher, many harmonics coincide, removing potential interactions. The simplest fraction is 1/2, an octave difference: all harmonics of the higher tone coincide with those of the lower, producing no more critical band interference than would occur between the harmonics of each tone heard separately.

When the two tones comprising the interval are sine waves, having no higher harmonics, the only interaction is between the two fundamentals. Simple-integer ratios should be irrelevant, in this case, and the only factor in the consonance or dissonance of the combination should be the difference between the frequencies; the smaller the gap, the more dissonantly they combine (unless they are very close, each *inside* the other's critical band, in which case they merge, and the interference is heard as an amplitude modulation imposed on a single tone).

In scaling studies previously conducted using intervals for stimuli, the consonance-dissonance dimension was expected to emerge as a major contributor to dissimilarity.

Levelt, van der Geer and Plomp [1966] applied the Method of Triadic Combinations to 15 intervals, with ratios ranging from 2:5 to 15:16. They pioneered the use of Balanced Incomplete Block Designs, applying a $\lambda = 4$ design. Eight subjects took part, each judging 35 triads (a $\lambda = 1$ block of the balanced design). Thus each block was judged by two subjects. Levelt *et al* collected two sets of data, the intervals being formed from simple sine tones in one set, and complex tones in the other, produced by filtering a train of short pulses with the right period (all harmonics approaching equal amplitude) to remove harmonics higher than 40 kHz.

To keep the average pitch of all intervals the same, the fundamental frequencies $f_1^{(l)}$ and $f_1^{(u)}$ of the lower and upper tones constituting each interval were adjusted: $(f_1^{(l)} + f_1^{(u)}) / 2 = 500$ Hz. Table 3.2 lists all $(f_1^{(l)}, f_1^{(u)})$ pairs. As the frequency ratio approaches unity, $f_1^{(l)}$ increases and $f_1^{(u)}$ decreases, and the frequency gap shrinks. It is convenient to express ratios and frequency gaps in semitones, a logarithmic scale. An Octave is 12 semitones. Two notes forming a musical Fourth, i.e. in a ratio of 2:3, are 5 semitones apart; the Fifth, with a 3:4 ratio, is 7 semitones.

This adjustment had the purpose of ridding the dissimilarities of extraneous, potentially confounding effects from pitch differences, but its efficacity is uncertain. In an experiment to be described below, scaling the same stimuli as Levelt *et al*, for most subjects the perceived pitch of an interval was governed, not by the (constant) average frequency, but by $f_1^{(l)}$ (or less often, $f_1^{(u)}$). Consequently, perceived pitch still varied between stimuli.

Table 3.2 $f_l^{(l)}$, $f_l^{(u)}$, frequency difference (all rounded to nearest Hz), and predicted dissonance (the logarithm of the lowest common harmonic), for 15 intervals

Interval	$f_l^{(l)}$	$f_l^{(u)}$	Δf_l	Dissonance	Interval	$f_l^{(l)}$	$f_l^{(u)}$	Δf_l	Dissonance
1:2	333	666	333	2.82	5:7	417	583	167	3.46
2:3	400	600	200	3.08	5:8	385	616	231	3,49
3:4	429	571	143	3.23	4:9	308	693	385	3.44
2:5	286	715	429	3.16	8:9	470	530	59	3.63
3:5	375	625	250	3.27	11:12	478	522	43	3.75
4:5	444	555	111	3.35	8:15	348	652	304	3.71
5:6	455	546	91	3.44	15:16	484	516	32	3.39
4:7	364	637	273	3.41					

Levelt *et al* arrived at a configuration in which the intervals are arranged in an approximate parabola, from large to small semitone difference. One interpretation of this arrangement is as a scale of increasing dissonance, but as we have seen, it could simply be that the stimuli were ordered by $f_l^{(l)}$. Their combined simple- and complex-tone data are scaled, in two dimensions, in figure 3.9.

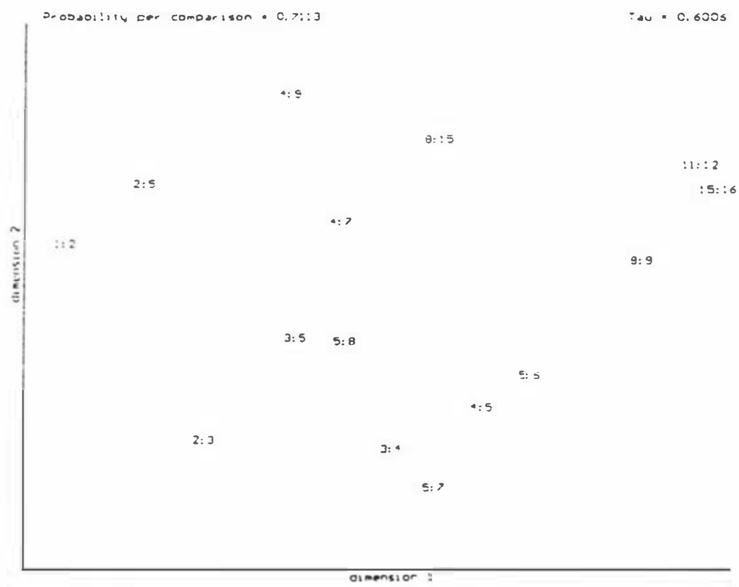


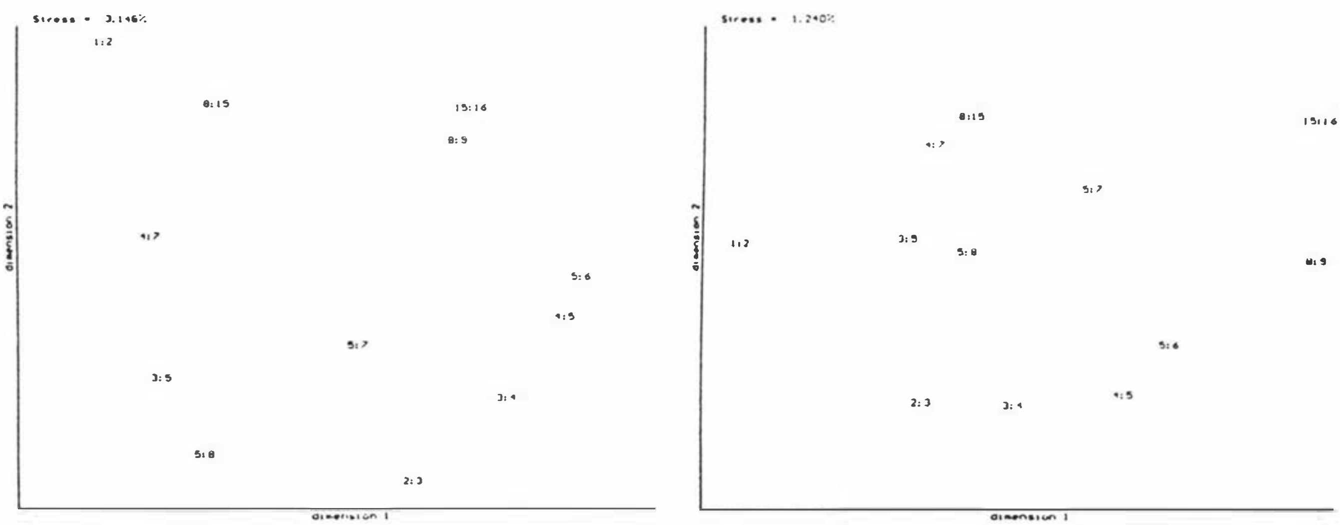
figure 3.9 Two-dimensional configuration for vote-counted triadic data for 15 intervals from Levelt *et al*

A second prominent aspect of their configuration is the second dimension: the parabolic sequence of intervals bends back on itself, to bring the extremes of frequency gap about as close to each other as both are to intermediate gaps. It may be that as well as pioneering the use of balanced designs, Levelt *et al* were early victims of the 'horseshoe artifact' [Shepard, 1974], exaggerated by vote-counting the triads before MDS.

But in this case, the horseshoe may reflect a genuine feature of the data. This is the special relationship enjoyed by pairs of intervals for which the sum of the semitone gaps is 12. For instance, the Fourth and the Fifth: if $f_l^{(l)}$ is in a ratio of 2:3 with $f_l^{(u)}$, then $f_l^{(u)}$ is in a ratio of 3:4 with ($2 f_l^{(l)}$). To an extent, $f_l^{(l)}$ and ($2 f_l^{(l)}$) are interchangeable in musical terms, and the intervals are correspondingly similar. The archetypal pair of similar intervals is of course 1:1 and 1:2, the Unison and the Octave, with semitone gaps of 0 and 12. 3:5 is close to 5:6, etc.

These similarities are hard to accommodate in a spatial model. The configuration must compromise between the close proximity of 3:5 to 5:6, and the remoteness of 3:5 from the neighbours of 5:6. In the event, it bends back on itself, the hinge being the six-semitone-gap interval (the Tritone, also known as the diminished Fifth or the augmented Fourth; a frequency ratio of 5:7). A second bend is added to the configuration, at the Octave, by the proximities of 2:5 to 4:5 and 5:8, and of 4:9 to 8:9.

Figure 3.10 Configurations for confusion data for 12 intervals from Plomp *et al* (a, left) and Killam *et al* (b, right)



For 12 of these 15 intervals, pairwise similarities have been measured in confusion experiments [Plomp, Wagenaar & Mimpen, 1973; Killam, Lorton & Schubert, 1975], in which an index of the similarity is how often one interval is misidentified as the other. The 15 subjects in Plomp *et al* were musically sophisticated (conservatory students), trained to concentrate on the key of notes as more important than octave differences between them; consequently, the 8:15 interval was frequently misidentified as 15:16 and *vice versa*, 5:8 and 4:5 and *vice versa*, and so on. In thir collected confusion matrices (one for simple-tone intervals and one for complex tones), this shows up as a band of high confusions along the minor diagonal, intersecting the band along the major diagonal consisting of confusions of intervals with their neighbours in the sequence. The result of applying MDS to the combined matrices is the two-dimensional configuration of figure 3.10(a) – again, a sequence curving back, with a special status for the Tritone.

In Killam *et al*, the subjects were less experienced (15 undergraduate music students). The stimuli were longer (0.1 and 0.2 seconds, instead of 0.015 to 0.125 s – mere blips – in Plomp *et al*), easing the identification of near-unity ratios, since there is more opportunity to observe the 'beating' between the fundamentals of small-frequency-gap intervals. One anticipates a more linear sequence, less curvature. Indeed, see figure 3.10(b). An added complication in Killam *et al* is that their published confusion matrix averages data for melodic as well as harmonic intervals.

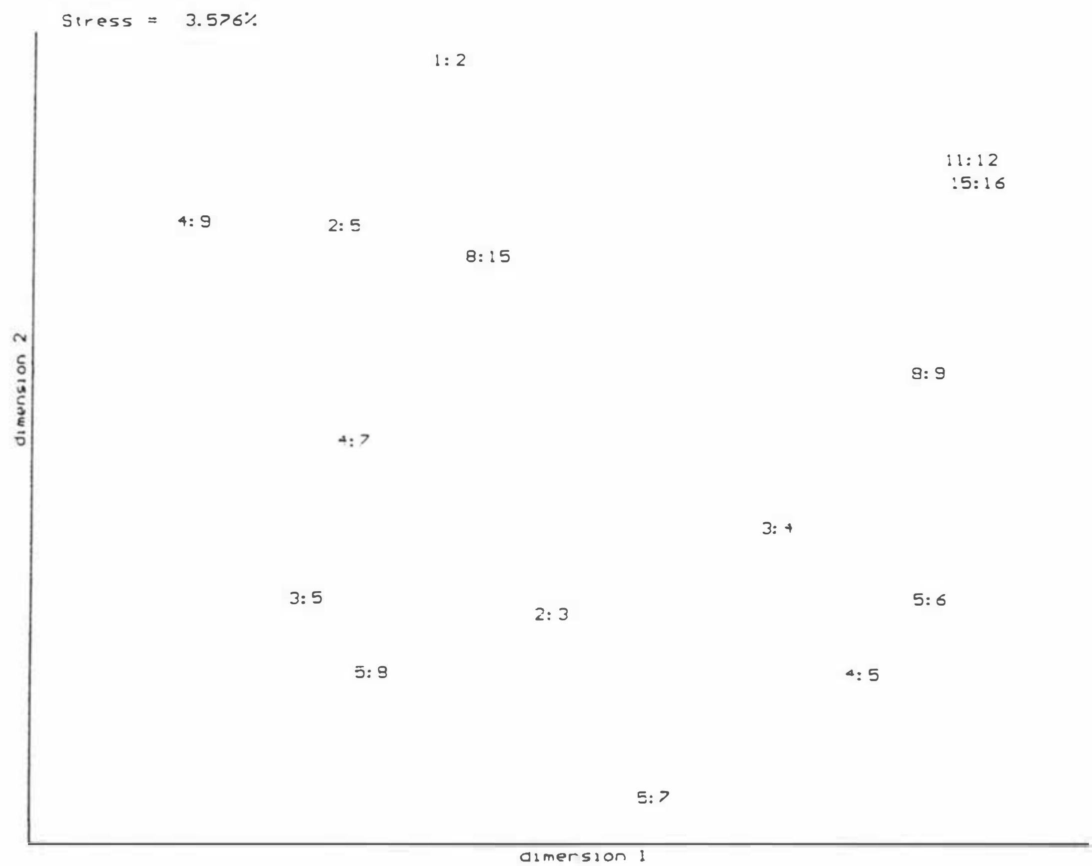
Balzano [1977] found the same circle of intervals, hinged at the Tritone, using as his index of similarity the times subjects took to arrive at same / different judgments. Finally, Shepard [1974] derived the same configuration from direct pairwise-comparison data.

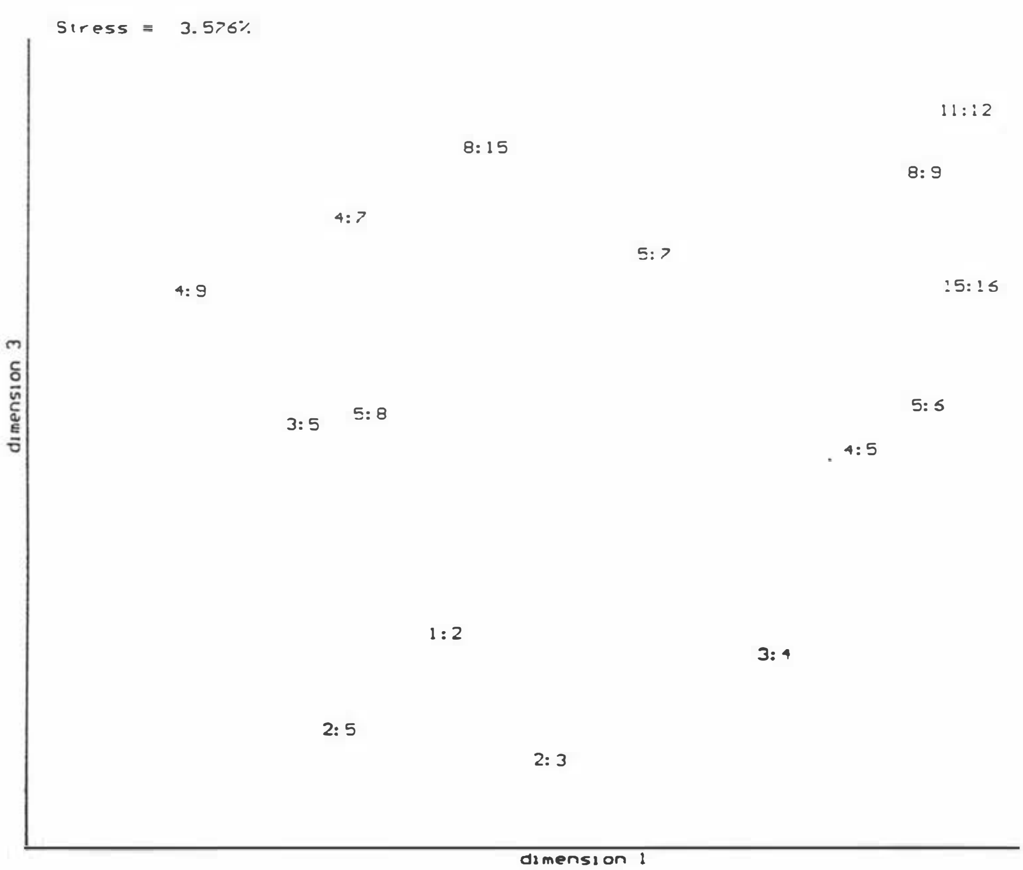
It isn't really important whether the proximity of extreme intervals in Levelt *et al* is genuine, or an artifact of the scaling process; in this context, what does matter is the way it obscures other contributions to dissimilarity, such as consonance / dissonance. As in the Shepard-tone experiment, the contributions of additional dimensions to inter-stimulus distance can only be approximated crudely, by displacing points radially towards or away from the centre of the circle. Otherwise a third dimension is required (or more).

I checked this point by jointly scaling the published data (Killam *et al*; Levelt *et al*; Plomp *et al*) in three dimensions. See figure 3.11. The stimuli are arranged along the third dimension from discordant intervals (8:15, 11:12, 15:16) at one end to euphonious ones (1:2, 2:3, 3:4) at the other. This leaves less extraneous variation to confound the neat frequency-gap sequence of the intervals in dimensions 1 and 2. Levelt *et al* deduced on theoretical grounds that the logarithm of Lowest Common Multiple($f_1^{(l)}$, $f_1^{(h)}$) should predict an interval's dissonance reasonably well. In this case the correlation between this dissonance index (last column in Table 3.2), and the positions of the intervals along the third dimension of figure 3.11, is 0.823.

Applying the INDSCAL model (described in Chapter 4) reveals that the matrices for complex-tone dissimilarities had higher saliences on the third dimension than those for simple tones, as one would expect, from the lack of higher-harmonic contributions to dissonance in the latter.

Figure 3.11 Dimensions 1 and 2 (upper), and 1 and 3 (lower) of a three-dimensional configuration for 15 intervals, combining data from Levelt *et al*, Plomp *et al* and Killam *et al*, rotated to maximise the correlation between the third dimension and the expected dissonance indices of Table 3.2





Experiments with Sounds (3)

I now describe an attempt to replicate the simple-tone interval part of Levelt *et al.* The experiment differed only in the details of which triads were presented to the subjects. It used primary-comparison triads, which in retrospect was not such a good idea, since for stimulus sets aligned along one dominant axis, as is the case here, secondary comparisons are useful for locating the stimuli upon minor axes.

Instead of a Balanced Incomplete Design, the experiment involved selecting the incomplete sets of triads interactively on the basis of previous triad responses. Details of Triskele, the program performing this selection, as well as presenting the triads, appear in Chapter 8. For now, suffice to say that a 'basis set' of stimuli is mapped, to start with, providing 'landmarks' which Triskele uses to triangulate the remaining stimuli. The idea behind this procedure is that when the number of triads is limited, each should be invested carefully, in the way that will return the greatest dividend in terms of information about the locations of the stimuli. Balanced Incomplete Designs are sub-optimal, since many of the comparisons which they call upon the subject to make have answers which merely confirm what is known from other triads.

I thought it best to test Triskele thoroughly before moving on to a study of baby cries.
11 subjects were recruited from amongst acquaintances. Each provided 168 triads, 84 for mapping a basis set of 9 stimuli, and seven triads for locating each of the remaining six stimuli. The membership of the basis set, and the sequence in which the non-basis stimuli were located, was different for each subject. The time taken to judge all the triads varied considerably, one subject accomplishing this feat in less than an hour, while another devoted three hours to the task, trying all possible combinations of the stimuli in each triad, playing and replaying them in a search for the sequence which most effectively emphasised the oddness of the odd-one-out.

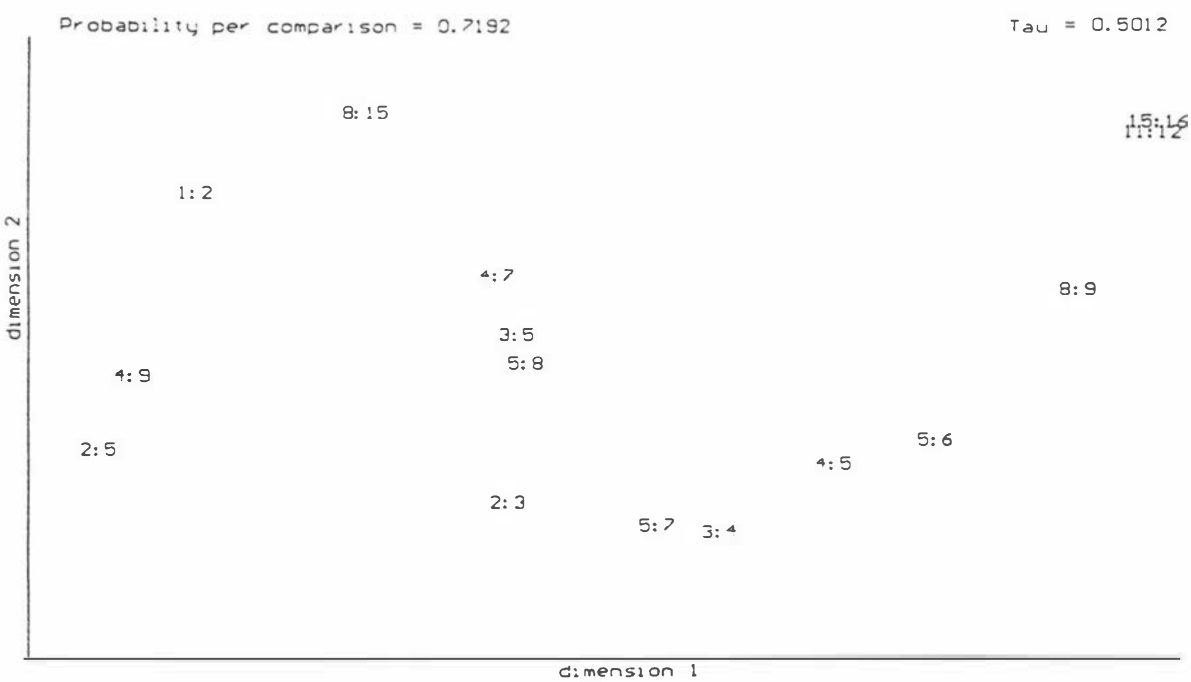
The responses from two subjects were barely discernable from chance. They did not seem to have grasped the nature of the task. These sets are omitted from the analysis.

The configurations from the remaining nine data sets are in general agreement with each other, and also with the published matrices discussed above (figure 3.12). The sequence of elements along the main axis (be it frequency gap or perceived pitch) is very clear, and that axis is bent in the same two points by the special proximity of interval pairs such as 5:8 and 4:5, requiring a second dimension. However, the degree of that bend is smaller than in Levelt *et al.* I remain suspicious that their parabolic configuration was exacerbated by vote-counting artifacts.

When the configuration is allowed to expand into a third dimension, the same segregation of the intervals by dissonance occurs.

This pilot study does not demonstrate any superiority of the Triskele procedure over conventional BIDs, since a BID with $\lambda = 4$, enough to provide a good recovery of the configuration, would involve the same number of triads. Further work is required.

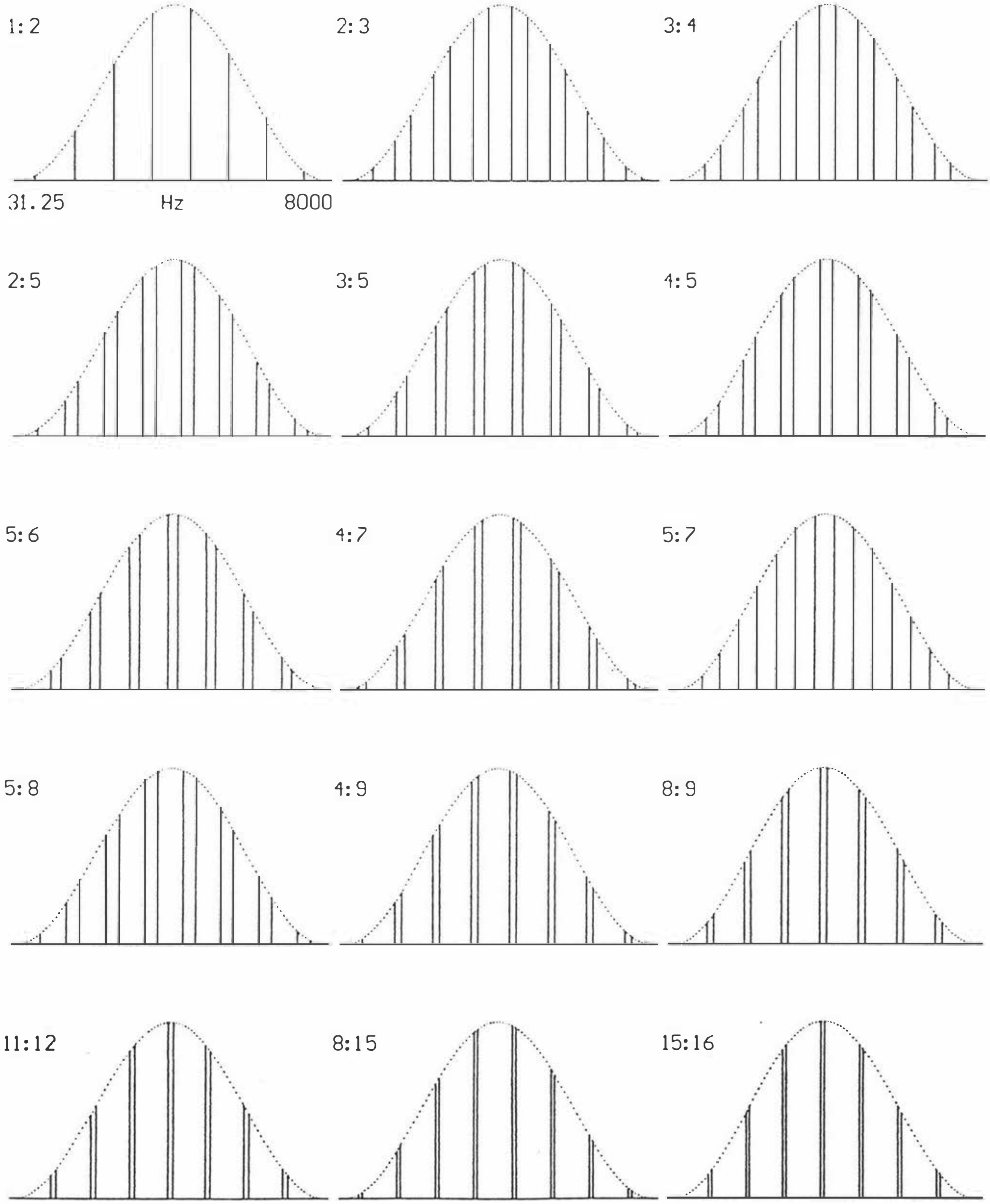
Figure 3.12 Two-dimensional configuration for 15 simple-tone intervals, using triadic data from nine subjects.



Experiments with sound (4)

In this fourth experiment, I applied the triadic method to complex intervals. In order to minimise differences of apparent pitch, and focus on differences of consonance or dissonance, the stimuli were synthesised by combining Shepard tones. I synthesised 15 Shepard tone intervals using the $f_l^{(l)}$ and $f_l^{(u)}$ from Levelt *et al.* f_p in this case was 500 Hz. Figure 3.13 plots the amplitudes of the components (vertical axis) against frequency (the horizontal axis, on a

Figure 3.13 Spectra of 15 Shepard-interval stimuli (frequency scales are logarithmic)

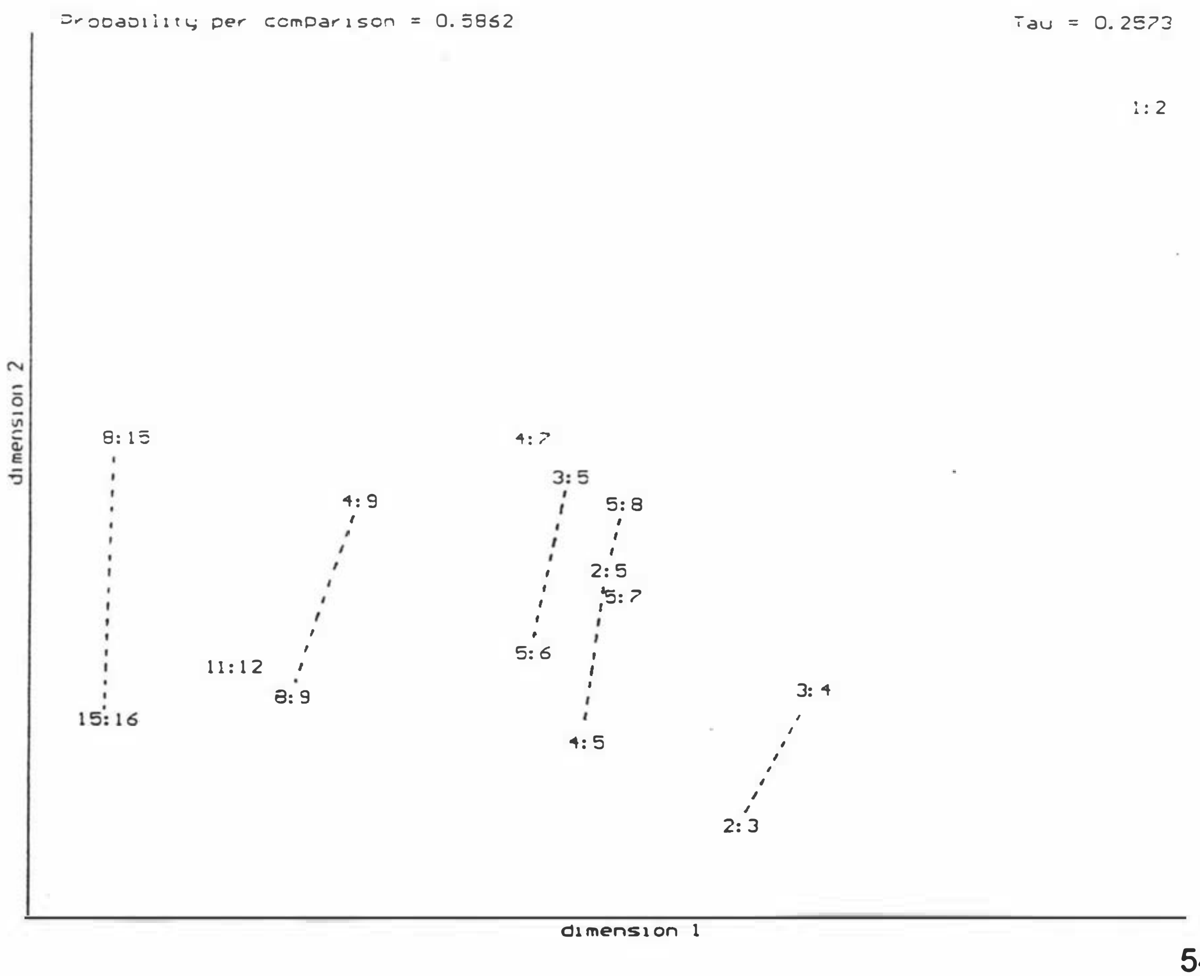


logarithmic scale) for these stimuli. Chords composed from Shepard tones have previously been used to study musical keys [Krumhansl, Bharucha & Kessler, 1982; Krumhansl & Kessler, 1981].

In these stimuli, the special relationship between pairs of intervals with semitone gaps summing to 12 is reinforced. If a stimulus contains $f_1^{(l)}$ and $f_1^{(u)}$ in the ratio 2:3, by definition it also contains ($2\,f_1^{(l)}$), so it is simultaneously a 3:4 interval. The stimuli labeled 2:5, 4:5, and 5:8 are effectively all the same interval (although they differ in the actual frequencies of their constituent tones). Figure 3.13 may clarify this. There is a limit, then, to how widely-separated the constituent tones can be in these Shepard intervals. Starting with the 15:16 stimulus, the gap between $f_1^{(l)}$ and $f_1^{(u)}$ can be progressively increased as far as the Tritone; increasing the gap any further brings $f_1^{(l)}$ closer to $f_1^{(u)}/2$.

As anticipated, a frequency-gap axis dominates the MDS solutions. Figure 3.14 combines data from nine subjects. Some differences in perceived frequency remain, hence the second dimension, serving to separate same-interval stimuli (linked by dotted lines in the diagram).

Figure 3.14 Two-dimensional configuration for 15 Shepard intervals using triadic data from nine subjects



More precisely, it separates intervals where the gap between the two highest-amplitude components is less than six semitones (e.g. 15:16, 5:6, 3:4) from those with a gap greater than six semitones (8:15, 3:5, 2:3). The INDSCAL model (Chapter 4) was used to find the best-fitting axes for this configuration.

Judging from its position in the configuration, the Octave stimulus is qualitatively different from the others. The octave has only half as many constituent tones, making it stand out, for most subjects, in most of the triads containing it, and placing it high on the second dimension.

One exception to the orderly arrangement of the intervals along a frequency-gap axis is the Tritone, 5:7. As we have seen, it should be the most harmonious of the stimuli, apart from the Octave: the particular harmonics which make a tritone combination of complex tones so dissonant are absent from these Shepard-tone intervals. However, it takes fourth place. This could reflect the unfamiliarity of the Tritone, which is avoided in Western music, whereas 2:3 and 3:4 ratios occur frequently. But the subjects were not particularly musically experienced. Alternatively, this observation could be taken as evidence for the 'periodicity' or 'long pattern' theory of human responses to tones: that the relatively long period that it takes a combination of incommensurate frequencies to repeat (as opposed to simple ratios like the Fourth, where the waveform of the combination has a frequency of $3\ f_l^{(l)} = 2\ f_l^{(u)}$) is perceived directly as dissonant, regardless of the presence or absence of higher harmonics.

Further experiments would be required to settle this point. One might, for instance, construct intervals using "stretched" tones [Mathews & Pierce, 1980; Pollack, 1978; Slaymaker, 1970] in which the ratio between successive "octave" harmonics (more correctly in this case, *partials*) is slightly greater than 2. If the ratio is not too great (e.g. 2.05), the ear still fuses such tones into a gestalt, though one having a percussive, bell-like timbre instead of the organ quality of a Shepard tone. Intervals formed from stretched tones should still be recognisable (with the Tritone still unfamiliar), but they would all have long periods.

Another possibility is to find subjects accustomed to non-Western styles of music.

The nine subjects were recruited opportunistically. This is the Casablanca technique: "Round up the usual suspects". Each listened to a different set of triads, the lists being BIDs with $\lambda = 3$ or 4

(105 or 140 triads), except for two subjects who did not complete the task. There are enough data to derive an individual configuration for each subject (apart from the two incomplete-data exceptions), though they show the same features as in figure 3.14. There are differences in dimensional salience. According to figure 3.15, the configuration for the author's data, I placed considerable emphasis on the consonance-dissonance dimension – hardly surprising, given my awareness of the purpose of the experiment. Another subject with high salience for dimension 1 had chosen to turn up the amplification of the stimuli unusually high. It is possible that at his preferred level of loudness, the dissonance of some intervals was increased by subjective harmonics and difference tones (generated by the non-linearity of the middle and inner ear).

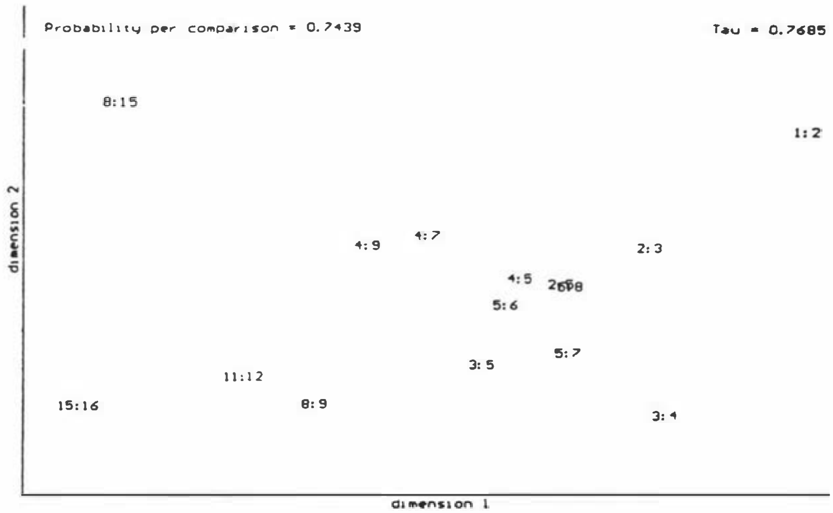
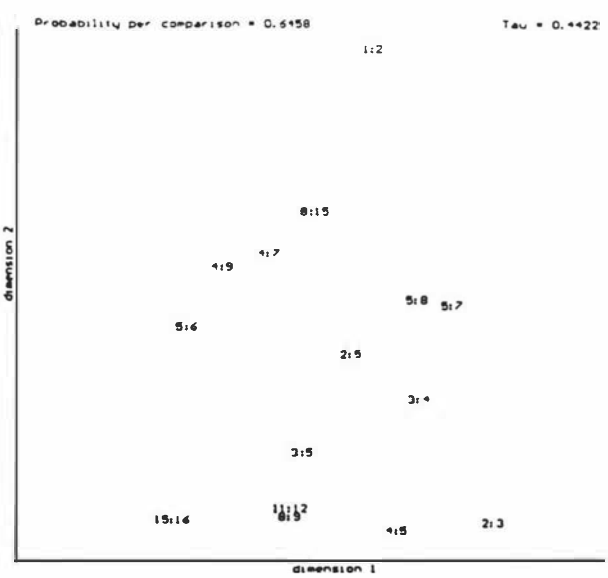


Figure 3.15 Configuration for subject DB (high salience on dimension 1)

For subjects AK, RX and JM, the residual differences in pitch were more salient: figure 3.16 is the configuration for AK.

Figure 3.16 Configurations for subject AK



Subjects were given no hints as to the criteria on which to judge "similarity". A tenth subject, a musician (or at least a guitarist) explained that his criterion for "similarity" between two intervals was how well they would go together in a chord. This is a non-geometrical model of proximity, the "ham and eggs" metaphor: ham and eggs are similar, because they go together well. That subject's results were too idiosyncratic to use in the pooled data.

4. GENERALISING THE MODEL

Introduction.

Several of the situations discussed in Chapter 3 involved pooling triads provided by more than one informant to make a single data set. The discussion assumed that subjects have similar response patterns, i.e. that their judgments are all based on copies of one mental map, and can safely be treated as if obtained from a single source.

This chapter abandons that assumption and considers general frameworks for coping with individual variations. It also introduces the concept of alternative definitions of "distance", other than the Euclidean distance function assumed in Chapter 2.

Though made in the context of triadic data, the generalisations are equally valid for other formats. See Wexler and Romney [1972] for a extraction of individual differences from specifically triadic data.

I have already asserted that a fully rank-ordered table of dissimilarities can be treated as pairwise comparisons. This point will be expanded on later; for now, note that multiple tables should be treated as multiple sets of pairwise comparisons. To turn dissimilarities into a matrix for MDS by averaging, as if they are ratio- or interval-level, is undesirable.

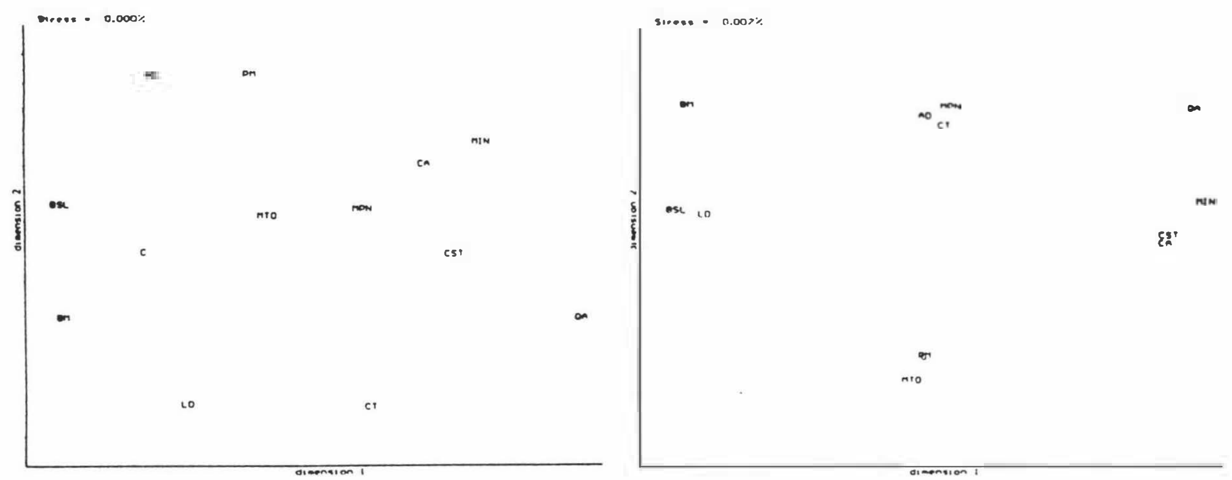
I will frequently refer to the spring metaphor of Chapter 2, to help maintain an intuitive feel for the models outlined in this chapter. The mathematical treatment remains elementary.

Divers Weights

"The difference between men is in their principle of association. Some men classify objects by color and size and other accidents of appearance; others by intrinsic likeness, or by the relation of cause and effect." [Emerson]

In some studies there are sufficient data to scale each subject's configuration in isolation, thus revealing the degree of variation that is concealed when they are combined. Figures 4.1(a), 4.1(b) are for two of the 47 subjects who contributed to figure 3.4. Recognising how subjects' maps differ is not a problem. If one pools the data, it is because individual deviations from the consensus mental map can be dismissed as random fluctuations, and eliminating them is desirable.

Figure 4.1 (a) and (b) Two sample individual configurations from the Coxon *et al* triad sets



An interesting situation, however, is where triads are shared out among subjects; there being too many for a single subject, with the result that no single subject provides enough responses to scale in isolation (the third-power proliferation can make this unavoidable). I seek a middle ground between analysis in isolation, and treating the data as if obtained from one subject; like drunks, unable to stand individually, the data sets should provide mutual support, while still allowing individual variations to emerge.

Even for complete data sets, one may want to perform some kind of mutually-constrained analysis, to separate genuine variations from the effects of random noise.

Permitting individual variations means that the dissimilarities perceived and compared by different subjects are not the same; each subject is working from a private mental map which is derived by some form of distortion from a group configuration. An obvious class of distortion is to vary the importance of the P dimensions between subjects. For two subjects S_1 and S_2, S_1 may find that elements' values along the first dimension make a greater contribution to inter-element dissimilarities, and hence to the selection of the odd-one-out from a triad, while S_2 is more sensitive or better attuned to the second dimension. L, the index of inter-subject variation extracted by Wexler and Romney [1972] from triads of kinship terms, in effect measured the relative salience between two dimensions (out of four).

This class of distortions is encompassed by the INDSCAL model [Carroll & Chang, 1970]. In this model, a modified Euclidean distance takes such variations into account:

$$d_{m,ij} = [\sum_{p=1}^{P} w_{mp} (x_{ip} - x_{jp})^2]^{1/2} \qquad (4.1)$$

$d_{m,ij}$ being the (i,j)-th dyad separation within the m-th subject's perceptual space; if M is the number of subjects, $1 \leq m \leq M$. The weights w_{mp}, representing the importance or salience of the p-th dimension to the m-th subject, form a M-by-P matrix W. One can interpret (4.1) by imagining the configuration X being copied from what Carroll and Chang call the "group stimulus space" into the subject's own perceptual space, suitably contracted or dilated:

$$x_{m,ip} = w_{mp}^{1/2}\, x_{ip}$$

and dissimilarities then computed with the standard Pythagorean formula.

For any X, a single point in configuration space, the range of private distorted versions X_m constitutes a $(P-1)$-dimensional subspace, and when X is varied in search of the optimal explanation of the observations, it is this whole *range* of configurations which vary; the parameters W, which specify which *members* of that range appply to each subject, should be optimised concurrently. As well as the group space, there is a P-dimensional "subject space", in which W is represented by the positions of M points with coordinates w_m. Optimising W consists of moving those points.

In this model, all subjects' dyad comparisons have a bearing on the group configuration, but the contribution of a given judgment depends on the judge's dimensional weights. Upgrading the spring model developed in Chapter 2, to include data from several subjects, is a matter of linking the nodes with multiple sets of springs, each set governed by the appropriate matrix of private distances D_m. For the sake of concreteness, suppose that S_1 and S_2 have both compared dyad (i,j) against (k,l), coming to the same conclusion, $\delta_{ij} > \delta_{kl}$. Even though they agree, their judgments

$$\delta_{1,ij} > \delta_{1,kl}$$
$$\delta_{2,ij} > \delta_{2,kl}$$

are kept separate, affecting only that subject's spring tensions. Let $\varepsilon_{m,ij,kl}$ be the value of $\varepsilon_{ij,kl}$ in subject m's data set:

$$\varepsilon_{m,ij,kl} = \begin{cases} 1 \text{ if } \delta_{m,ij} > \delta_{m,kl} \\ 0 \text{ otherwise.} \end{cases}$$

Contradictions between reconstructed distances and comparisons can now be summed over M subjects to calculate the Stress:

$$\text{raw S}(X, W) = \sum_{m=1}^{M} \text{S}_m(X, w_m)$$

$$\text{where S}_m(X, w_m) = \sum_{(i,j)\,(k,l)} \varepsilon_{m,ij,kl}\, \text{H}(d_{m,kl} - d_{m,ij})\,(d_{m,kl} - d_{m,ij})^2 \qquad (4.2)$$

For the simple cases of $M = 1$ (a single subject), or when data sets are pooled under the assumption that individual perceptions are the same (i.e. treated as replications), clearly

$$w_{mp} = 1 \text{ for all } m,p.$$

In other cases, add constraints on the weights, since it is undesirable for perceptual spaces to vary freely in scale between subjects:

$$\sum_{p=1}^{P} w_{mp} = P \text{ for all } m \qquad (4.3)$$

or equivalently,

$$w_{mP} = P - \sum_{p=1}^{P-1} w_{mp}$$

i.e. the weights vector w_m for the m-th subject has only (P-1) independent components. (4.3) restricts the corresponding point in subject space to lie on a (P-1)-dimensional hyperplane, intersecting the axes at (P,0,0,...0), (0,P,0,...,0),... (0,0,0,...,P). The $M = 1$ point (1,1,1,...,1) is also on this plane.

Further constraints can be imposed on W. We wish the group stimulus space to be at the centre of the individual spaces, i.e. the M points in subject space should have their centre of gravity at the $M = 1$ point, the "centroid configuration" [Lingoes & Borg, 1978]:

$$1 / M \sum_{m} w_{mp} = 1 \qquad (4.4)$$

Thus $1 / M \sum_{m} x_{m,ip}^2 = x_{ip}^2$.

W has only (M-1) (P-1) degrees of freedom.

As in Chapter 2, it remains to vary the sundry parameters – W as well as X – until $S(X,W)$ is minimal. The matrix decomposition used by Carroll and Chang [1970] assumes ratio data and is not applicable here. The absence of a matrix Δ of disparities precludes a matrix solution for W, as in ALSCAL [Takane *et al*, 1977]. Rather than Carroll and Chang's approach of successive approximation, we optimise the configuration progressively through the iterations of a step-wise hill descent. Each iteration replaces $X^{(t)}$ with $X^{(t+1)} = X^{(t)}$ + (step-size) ∇X, as before, also replacing $W^{(t)}$ with $W^{(t+1)} \;=\; W^{(t)} + \Delta W \;=\; W^{(t)}$ + (step-size) ∇W

The gradients are $\partial S(X,W) / \partial x_{ip} = \sum_m \partial S_m(X, w_m) / \partial x_{ip}$

where $\partial S_m(X, w_m) / \partial x_{ip} = \sum_j (\partial S_m / \partial d_{m.ij}) (\partial d_{m.ij} / \partial x_{ip})$

$$= -w_{mp} \sum_j (x_{ip} - x_{jp}) / d_{m.ij} \left(2 \sum_{(k,l)} \varepsilon_{m.ij.kl}\, \Theta(d_{m.kl} - d_{m.ij})\right) \qquad (4.5)$$

Similarly, $\partial S(X,W) / \partial w_{mp} = \partial S_m(X, w_m) / \partial w_{mp}$

$$= -1/2 \sum_{(i,j)} (x_{ip} - x_{jp})^2 / d_{m.ij} \left(2 \sum_{(k,l)} \varepsilon_{m.ij.kl}\, \Theta(d_{m.kl} - d_{m.ij})\right) \qquad (4.6)$$

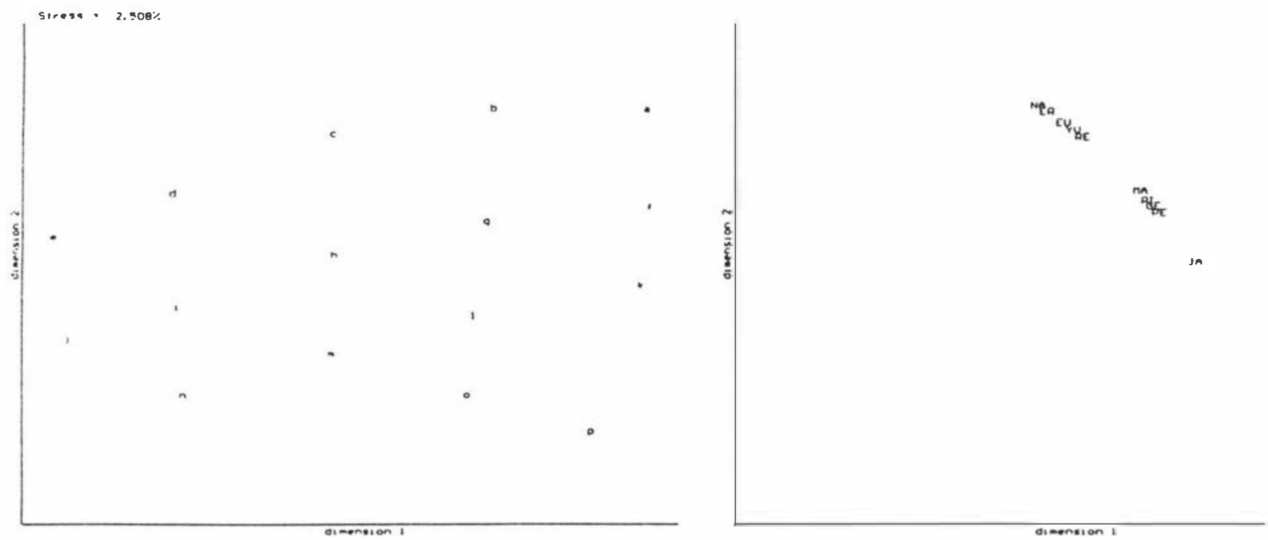
These calculations are easily implemented. MTRIAD calculates the personalised distances $d_{m.ij}$ for each subject, then iterates over all dyads (i,j) and (k,l) for which $\varepsilon_{m.ij.kl} > 0$, accumulating contributions to that subject's "spring forces". As well as contributing to Δx_i and Δx_j, the (i,j)-th spring force $f_{m.ij}$ contributes to Δw_{mp}, in proportion to the separation of i and j along the p-th axis.*

Consider ∇X_m (the contribution to ∇X from the m-th data set) in the form of a vector in configuration space. (4.6) amounts to projecting ∇X_m onto the subspace of allowable transformations of X. The component of ∇X_m tangential to that subspace is an indication of how much w_m should be altered. If ∇X itself has a non-zero projection onto the subspace, X needs to be stretched or compressed.

Like overall scale, the constraints (4.3) and (4.4) are enforced by rescaling after each ΔW.

*** By inspection of (4.5) and (4.6), the derivatives of S are defined for all X, W, although they are discontinuous, because of the step functions. Loss functions defined in terms of disparities [Kruskal, 1964] are not guaranteed to be everywhere differentiable [de Leeuw, 1988]. De Leeuw also discusses convergence properties when loss functions are minimised by using the majorization algorithm.**

Figure 4.2 Group configuration (a, left) and subject space (b, right) for Stalmeier-de Weert colour data (M = 10)



Applying this to data collected by Stalmeier and de Weert [1991b], with N = 16 colour stimuli and M= 10, in two dimensions, yields figure 4.2. Panel 4.2(a) shows the group configuration; panel (b) is the subject space, plotting $w_{m1}^{1/2}$ horizonally and $w_{m2}^{1/2}$ vertically. Individual variations are indicated; for instance, subject PE is evidently relatively attentive to the red-green axis. Comparing back to PE's individual configuration (figure 3.5) confirms this. But note the slight rotation between the two configurations. This highlights the important point that orientation in Euclidean space is arbitrary; rotation of the set of axes into another set leaves distances (and Stress) unchanged. The axes in 3.5 are inherited from the original metric solution for $X^{(0)}$. Principal Coordinates Analysis aligns the axes with any elongation of the configuration, so to maximise the variance they account for. Depending on the configuration, they may coincide with meaningful dimensions (as in this case), but one cannot rely on this.

Weighted Euclidean space – the "INDSCAL model" – breaks the symmetry. If the model is applicable, the configuration rotates in the course of minimising Stress, finding the alignment where the spread of subjects' weights is greatest. A sufficient reduction in Stress gives one confidence that the resulting axes are interpretable, as ways in which subjects can vary, if nothing else. This is the justification for using INDSCAL. The option for individualised weights is to be used cautiously, as with any model introducing additional degrees of freedom.[1]

[1] "Divers weights, and divers measures, both of them are alike, abominations to the Lord." [Proverbs 20:10]

We are fortunate in having external evidence to ease any doubts about figure 4.2. Of those ten informants, eight participated in a second experiment [Stalmeier & de Weert, 1991b], using a different procedure, with six fewer colours in the stimulus set ($N = 10$). Instead of the "Star-of-David" pattern, coloured circles (each subtending a visual angle of 1.06°) were presented in a triangle (with a 0.6° gap between each pair): subjects assessed the colours' dissimilarities consciously rather than relying on gestalt fusion. Analysing the results shows the same ordering of subjects. Moreover, these dimensional weights emerge when each of the four replications comprising each subject's data set is analysed separately.

Similar variations in colour sensitivity appear in previous studies: see Wish and Carroll [1974] for INDSCAL reanalyses of data collected by Helm, and by Indow and Kanazawa. Stalmeier [personal communication] suggests several explanations for these idiosyncracies. As well as differing cone-cell responses, affecting the relative strengths of the red-green and blue-yellow opponent processes, there is pre-retinal filtering of light, varying with the amount of (yellow) foveal macular pigment, and the extent of yellowing of the lens: also likely to align the second INDSCAL dimension with the yellow-blue axis (tritanopic confusion lines).

The POOC triadic data, for 13 names of occupations [Coxon *et al*, 1975] have already been mentioned (figure 3.4). Here $M = 47$. A Balanced Incomplete Design was used, with $\lambda = 2$, collecting only 52 triads from each subject, though not all subjects completed the whole list (in the despairing words of the User's Guide, "Missing data just isn't there"). Repeating the analysis with the INDSCAL model results in a reduced Stress. The evidence for the validity of these axes is that if the configuration is rotated, it rotates back to this preferred alignment. They concur with the dimensions proposed by Coxon and Jones, one being Income / Status, and the other corresponding to the amount of personal interaction involved in different occupations. The anticipated variation of weights with informant's social status did not show up.

When the model is applied to the musical-intervals data discussed in Chapter 3, combining all 14 data sets (three confusion matrices, two matrices of vote-counted triads, and nine sets of simple-tone triads from the third experiment), it turns out that six sets place high weights on the third dimension (simple-ratio consonance-dissonance). Three of these are the observations made using complex tones: two complex-tone confusion matrices [Killam *et al*, 1975; Plomp, Wagenaar, & Mimpen, 1973], and the vote-counted complex-tone triads [Levelt *et al*, 1966].

This fits in with the theory that simple ratios become a factor in consonance when intervals include harmonics. There are also three sets of triads having a high w_{m3}. One of these came from the experimenter, who was aware of the purpose of the study, while for a second set the subject was a musician.

Applied to the data from the 4th experiment (Shepard-interval triads), the INDSCAL model highlights two subjects placing a high weight on the consonance-dissonance dimension (w_{m1}, in this case). One was myself. The second, subject AG, chose to play the sounds loudly enough that subjective harmonics (produced by nonlinearities in the inner ear) were possibly contributing to dissonance.

A similar reduction of Stress with axial realignment occurs when one applies the INDSCAL model to the cry data of Lyons *et al* [1991], where $M = 8$. The result, Figure 4.3, confirms that the configuration derived earlier from subject B's data is representative of the other subjects.

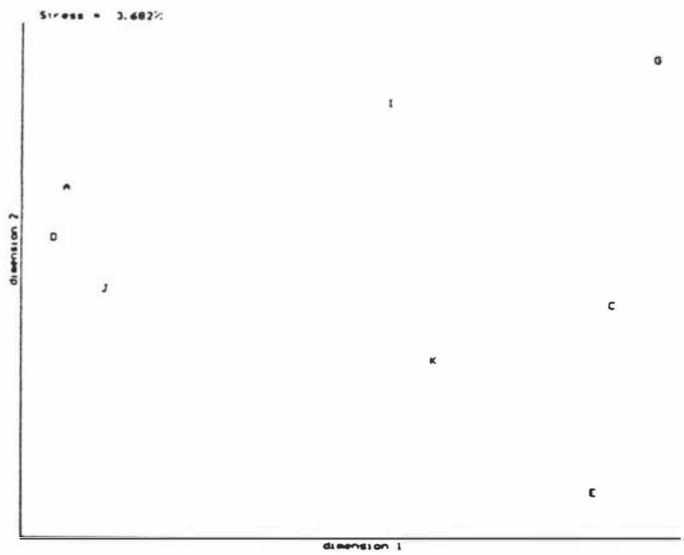


Figure 4.3 INDSCAL solution for cry triads

Sparse data sets

The $N = 16$ Stalmeier-de Weert material affords a demonstration that individual variations can still be discerned in sparser data sets. I constructed a new data set for eight of the 10 subjects, discarding 7/8 of the data and apportioning triads so that each triad was judged by one and only one subject (thus each data set comprises 70 triads). Using the MTRIAD and the vote-count procedures to recover the configuration for subject PE results in figures 4.4(a) and 4.4(b) respectively, and highlights the weakness of the latter when dealing with incomplete data. Moreover, the same individual variations still show up upon pooling the data sets (figure 4.5).

figure 4.4 MTRIAD (a, left) and vote-count (b, right) solutions for 1/8 of the Stalmeier-de Weert data for PE

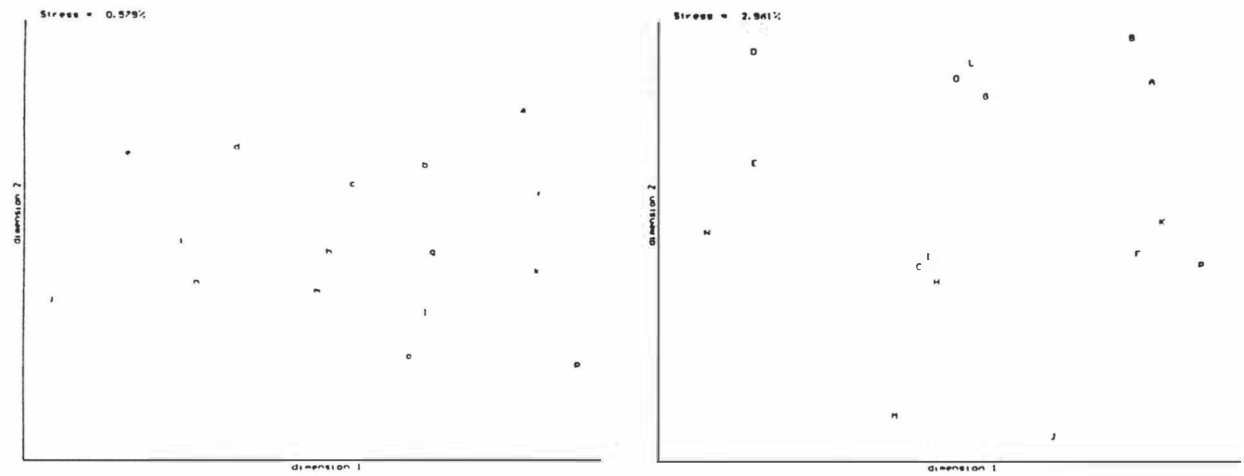


figure 4.5 Subject space for reduced data set I

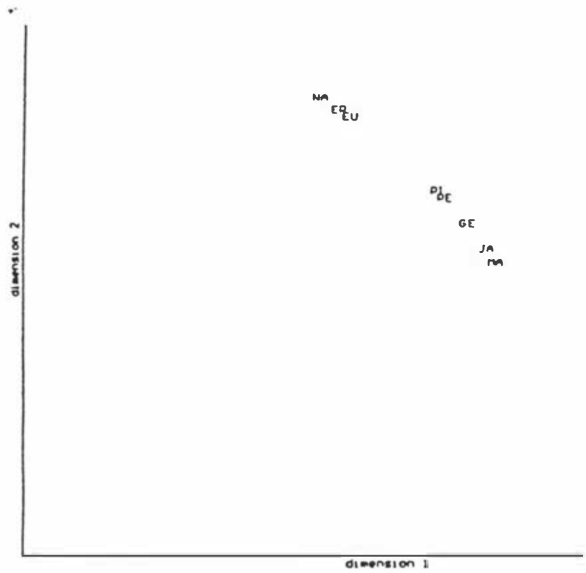
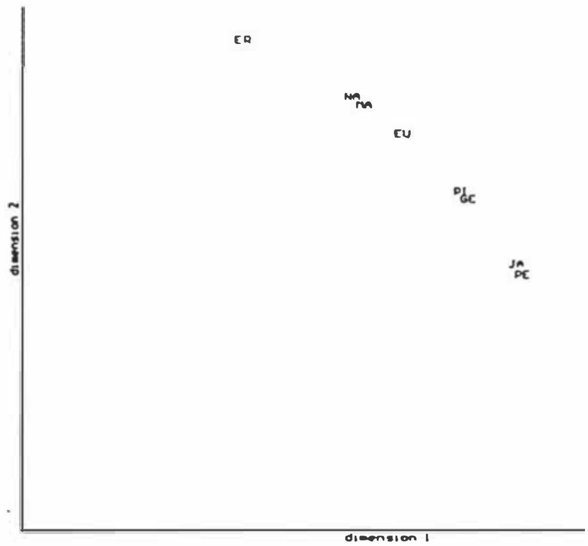


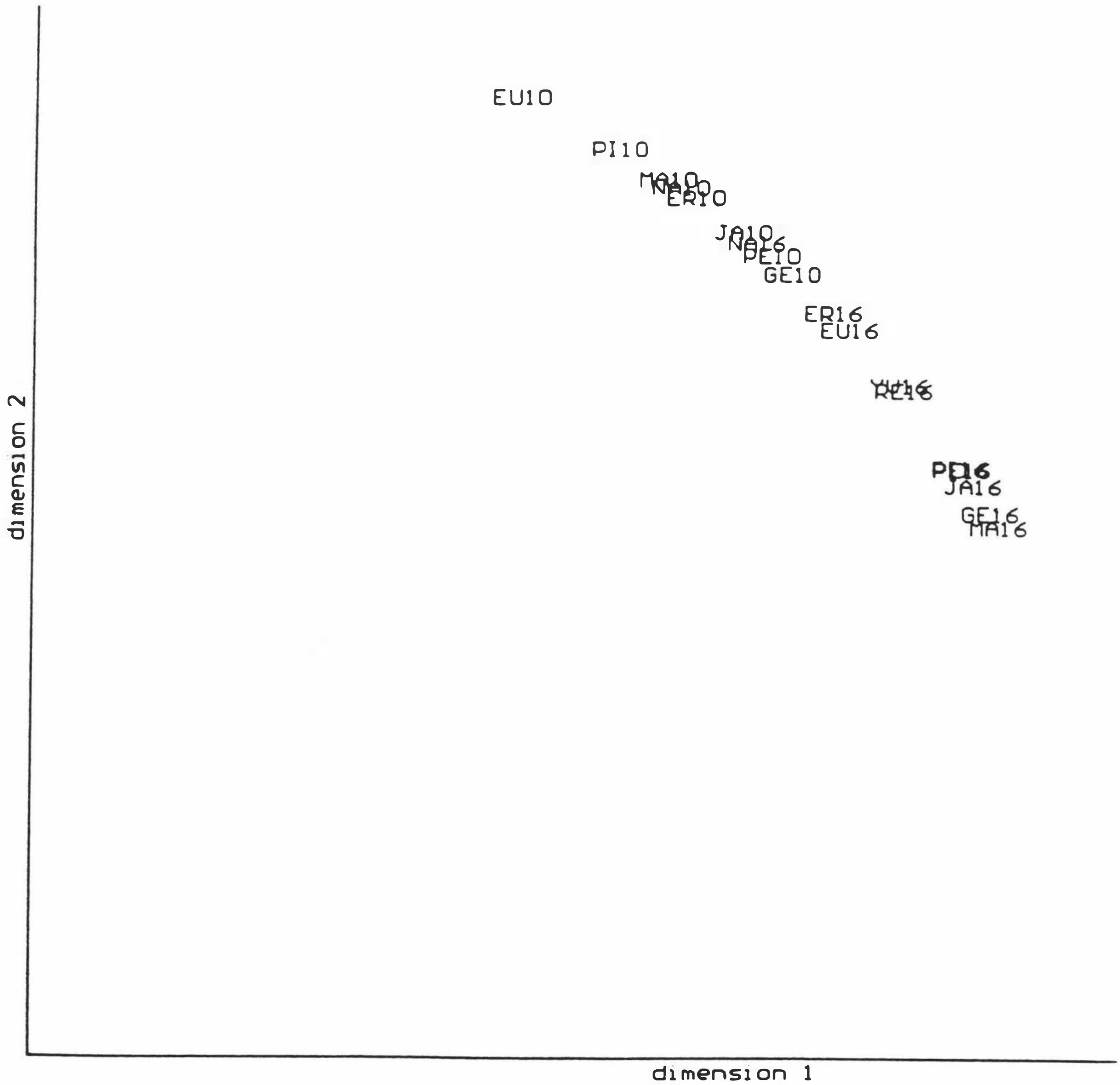
figure 4.6 Subject space for reduced data set II



A second demonstration apportions triads with some overlap. I winnowed subject ER's data to retain only those 210 triads <*i,j,k*> in which stimulus *j* is one of the colours 'a' and 'i'; subject EU retained only triads in which *j* is 'b' or 'j'; and so on to subject PI (the 210 triads in which *j* is 'h' or 'p'). The resulting data sets are conditional ranking data and must be combined for analysis; they are too imbalanced for treatment in isolation. The resulting subject space (figure 4.6) illustrates how individual variations are still resolved, even under these adverse conditions.

One point remains to be made about the Stalmeier-de Weert data. The set of 16 stimuli used in their first experiment [1991a] includes the 10 colours of the second experiment [1991b], making it possible to scale and weight them all in one grand combination. Figure 4.7 shows the unexpected result. Individual variations in colour sensitivity are superimposed on a larger difference between experiment designs. It would appear that separation along the Blue-Yellow dimension contributes more to the conscious assessment of colour dissimilarity, than to the mechanisms of gestalt fusion underlying the Star-of-David design. In retrospect, this should not have been so surprising, for edge distinctness presumably plays a part in gestalt fusion, and colour awareness and edge detection are known to be mediated by separate cellular pathways in the retina and brain, with the latter displaying a partial tritanopic colour-blindness [Livingstone, 1988]. Edge distinctness is not a valid index of colour dissimilarity. Scaling subjects' assessments of the distinctness of edges between monochromatic stimuli leads to a colour ellipse instead of the expected colour circle [Tansley & Boynton, 1976].

Figure 4.7 Subject space, two different protocols, $N = 16$ and 10



The observation demonstrates one situation when the INDSCAL model serves a useful purpose: when one suspects that multiple data sets originate in distinct sub-populations, differing in their perceptual maps, and one seeks to resolve them. This includes issues such as diagnosing colour-vision deficiencies.

Sometimes the results of a course of tuition are also measurable as dimensional weights, as in studies like Pollard-Gott [1983] (where subjects comparing sections of music for dissimilarity showed growing awareness of a "thematic" dimension as exposure increased, with musicians weighting the thematic dimension most highly), and Green, Jones and Gustafson [1987].

In an application of the INDSCAL model to study pain [Clark, Janal, & Carroll, 1989], 48 subjects (24 healthy volunteers, and 24 cancer pain patients) provided dissimilarities between 9 pain descriptors. The "pain intensity" dimension of the resulting configuration was more salient for the patient group; healthy subjects were more attentive to the "emotional quality" dimension of the descriptors. The authors concluded: "The MDS approach suggests that questions such as, 'Is this patient really in pain or merely depressed?' will be replaced by 'What are the patient's coordinates in the multi-dimensional global pain space and what dimensions are most relevant for him?'" (page 297).

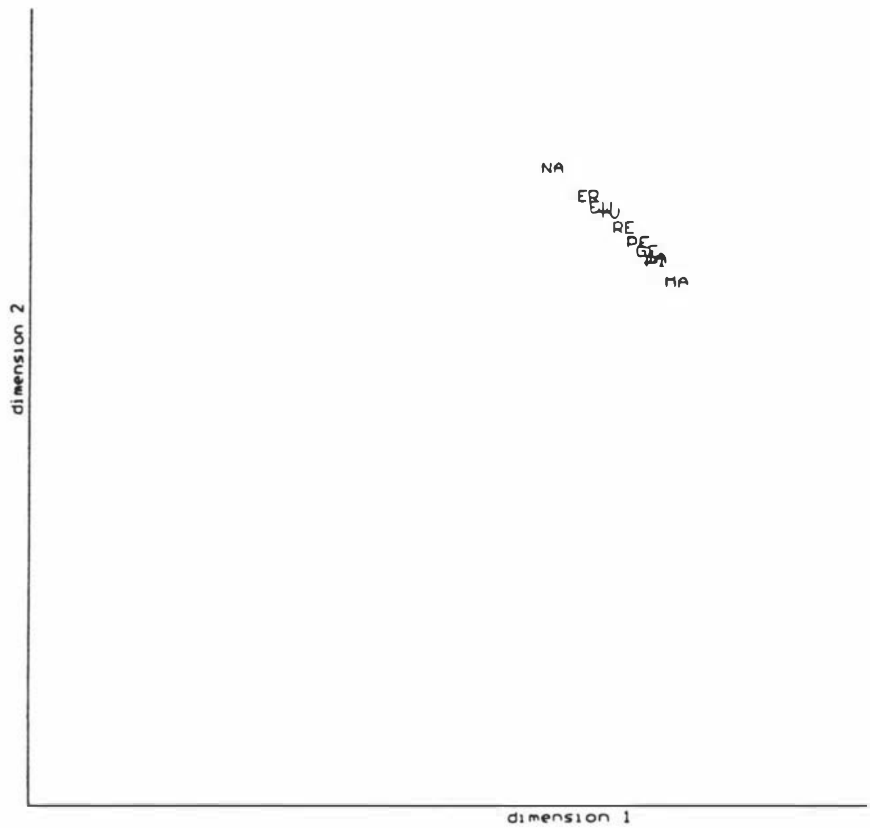
When the configuration is known in advance, and need not be recovered from the data at hand, a complete set of dissimilarity comparisons can be replaced with comparisons selected solely to determine W. It is clear from (4.4) that $\partial S_m / \partial w_{mp}$ is more sensitive to some comparisons than others. When $S(X, W)$ is minimised by varying W (i.e. holding X constant), these are the comparisons with the greatest potential effect on w_m. The relative salience of axis p is most sensitive to comparisons between two distances which are roughly equal (in the stimulus space for the m-th subject), with one dyad parallel to the axis and the other perpendicular to it.

By now, no-one will be surprised when I simulate an example of W-only scaling using the Stalmeier-de Weert data ($N = 16$). From each subjects' data set I took their responses to 16 triadic comparisons (listed in Table 4.1), picked for their weight-discriminating value by examining figure 4.2(a). Scaling them under the INDSCAL model, with X held constant by including a table of distances d_{ij} in the analysis (these being the inter-point distances obtained by analysing the complete data), results in figure 4.8, in good agreement with figure 4.2(b).

Table 4.1. 16 triadic weight-discriminating comparisons.

i	*j*	*k*
c	a	p
d	b	o
a	c	o
c	d	n
d	e	j
a	g	o
c	h	g
d	i	j
a	k	o
b	l	i
c	l	k
c	m	j
c	o	n
b	o	i
a	p	m
b	p	n

Figure 4.8 Subject space with X constant and 16 comparisons per subject



If this were used as a test of colour perception in practice, where the range of potential variation is wider, a larger set of triads would be required. Two students at Massey University (Emma Barraclough and Don Kirkland) are conducting research along these lines.

The Star-of-David experiment design involved in acquiring this data has much in common with the widely-used Ishihara test for colour-vision deficiency. Both are gestalt procedures, which present the visual system with two alternative interpretations, mutually incompatible, to be resol ved by vision's opponent processes. The interpretation which pairs up closer colours will "gel" and exclude the alternative interpretation. Ignoring incidental details of gestalt design, the major contrast between the two is the use of a white background in the Ishihara test, preventing edge-distinctness effects from confounding the outcome. I conclude that Ishihara results are triadic data.

Consider the Farnsworth-Munsell D15 test of colour vision in the same light. This involves two sets of 16 colour samples, each set varying only in hue, spaced at equal intervals around a colour circle (colour brightness and saturation are fixed at 5/4 for one set and 8/2 for the second, so there are separate circles, one for each set). With one sample serving as the initial anchor, the subject chooses the colour closest to it. That choice becomes a new anchor and the closest of the remaining colours to it is chosen; a process which continues until the 16 colours are all arranged in a sequence. This is still a form of dissimilarity comparison (specifically, it is a pick $1/N$ procedure – to be discussed in Chapter 6 – modified by the dwindling choice for "closest colour"). For a subject whose stimulus space is sufficiently distorted, with the colour circles collapsed into ellipses by extreme values for w_{mp}, the sequence is abnormal.

This procedure appears in a more extensive form as the Farnsworth-Munsell 100-Hue test.

The IDIOSCAL model

Not all forms of variation between individuals' perceptual maps (or between the same individual under different conditions, or at different stages of training) lend themselves to description in terms of dimensional weights. One step up in generality is the IDIOSCAL model [Carroll & Wish, 1974], which still involves compression or expansion of the group configuration along a set of dimensions, but these are not necessarily the axes of the stimulus space. One form of this model (the "Carroll-Chang decomposition") subjects the configuration to a rotation before weighting is applied[2]. This introduces another $M(P-1)$ degrees of freedom since a separate rotation is fitted for each subject. The classification of colour vision deficiencies is a situation to which this model is admirably suited. Depending on which of the three cone-cell pigments is abnormal, the private colour space (in two dimensions) of a colour-vision-deficient subject is compressed along one of three axes. Two of these axes (for proteranopes and deuteranopes) are only 30° apart, with roughly 90° between them and the third, tritanope axis, so the INDSCAL model is a useful first approximation.

In the IDIOSCAL model, as well as looking for a component of ∇_m indicating that X_m should be stretched along the m-th set of axes, we look for a rotational component. The range of transformations X_m of a given X now constitute a $2(P-1)$-dimensional subspace of configuration space, but finding the individual-variation parameters is still a matter of projecting ∇_m onto the subspace to extract its tangential component.

[2] There is also the Tucker-Harshman decomposition, also known as the PARAFAC model, which allows for affine transformations of the group configuration: the axes of expansion or compression are not constrained to be at right angles.

So far I have used W to label the M-by-P matrix of dimensional weights, with individual subject weights being w_{mp}, or w_m to label a point in subject space. Often in scaling theory, weights are represented as P-by-P diagonal matrices $\mathbf{W}_m$, while the coordinates of the i-th point, in the personal and group stimulus spaces respectively, are written as column vectors $\mathbf{x}_{m,i}$ and $\mathbf{x}_i$. This has the advantage of letting one write $\mathbf{x}_{m,i} = \mathbf{W}_m \cdot \mathbf{x}_i$ (and $\mathbf{X}_m = \mathbf{W}_m \cdot \mathbf{X}$).

The IDIOSCAL model interpolates a rotation matrix, $\mathbf{R}_m$:

$$\mathbf{X}_m = \mathbf{W}_m \cdot \mathbf{R}_m \cdot \mathbf{X}$$

So $$d_{m,ij}^{\ 2} = (\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}} \cdot \mathbf{R}_m^{\ \mathrm{T}} \cdot \mathbf{W}_m^{\ \mathrm{T}} \cdot \mathbf{W}_m \cdot \mathbf{R}_m \cdot (\mathbf{x}_i - \mathbf{x}_j) \tag{4.7}$$

In two dimensions, as well as a single weight parameter w_{m1} (since $w_{m2} = 2 - w_{m1}$), there is a single rotation parameter per subject, θ_m, and

$$\mathbf{R}_m = \begin{bmatrix} c_m & -s_m \\ s_m & c_m \end{bmatrix} \qquad \text{where } c_m = \cos\theta_m,\ s_m = \sin\theta_m$$

As before, the situations we are interested in involve no distance estimates δ_{ij}, and (4.7) cannot be decomposed directly. Once again, the answer is to minimise Stress iteratively. Inserting (4.7) into (2.9), the differentiation is not particularly laborious. We get

$$\partial S_m / \partial x_{i1} = \sum_j f_{m,ij} \left[(w_{m1} c_m^{\ 2} + w_{m2} s_m^{\ 2})(x_{i1} - x_{j1}) + (w_{m2} - w_{m1})\, c_m s_m (x_{i2} - x_{j2}) \right] / d_{m,ij}$$

$$\text{where } f_{m,ij} = 2 \sum_{(k,l)} \varepsilon_{m,ij,kl}\ \Theta(d_{m,kl} - d_{m,ij})$$

$$\partial S_m / \partial x_{i2} = \sum_j f_{m,ij} \left[(w_{m1} s_m^{\ 2} + w_{m2} c_m^{\ 2})(x_{i2} - x_{j2}) + (w_{m2} - w_{m1})\, c_m s_m (x_{i1} - x_{j1}) \right] / d_{m,ij}$$

$$\partial S_m / \partial w_{m1} = \sum_{(i,j)} f_{m,ij} \left[c_m (x_{i1} - x_{j1}) - s_m (x_{i2} - x_{j2}) \right]^2 / d_{m,ij}$$

$$\partial S_m / \partial w_{m2} = \sum_{(i,j)} f_{m,ij} \left[c_m (x_{i2} - x_{j2}) + s_m (x_{i1} - x_{j1}) \right]^2 / d_{m,ij}$$

$$\partial S_m / \partial \theta_m = \sum_{(i,j)} f_{m,ij} (w_{m2} - w_{m1}) \left[c_m (x_{i1} - x_{j1}) - s_m (x_{i2} - x_{j2}) \right] \left[c_m (x_{i2} - x_{j2}) + s_m (x_{i1} - x_{j1}) \right] / d_{m,ij}$$

Figure 4.9 was produced by applying this model to four sets of distance-comparison data, all elicited using the D15 stimuli [Barraclough, unpublished data]. Two data sets used the standard D15 protocol, for the saturated and unsaturated stimulus sets (12 and 11 subjects respectively). The other two used a 'hierarchical sorting' procedure, which I describe in Chapter 7, performed on both stimulus sets (by 19 subjects for the saturated, and 17 for the unsaturated sets). I do not show the configuration itself (which was circular, by constraint). Instead, figure 4.9 is a depiction of subject space. It uses radial coordinates, plotting $(w_{m1}-1)^{1/2}$ against (2θ). This is to eliminate a redundancy in the parameters of the IDIOSCAL model. A rotation through θ, followed by dimensional weighting with weights $(w_{m1}^{1/2}, (2-w_{m1})^{1/2})$ has the same effect as rotating through $\theta + 90°$ and weighting with $((2-w_{m1})^{1/2}, w_{m1}^{1/2})$.

Subjects who had been diagnosed as colour-vision deficient using the Ishihara test or anomaloscope examination are marked on the diagram. There is a clear separation of the protanope and deutanope groups. Had there been tritanopes among the subjects, they would presumably have been positioned somewhere to the left of the diagram.

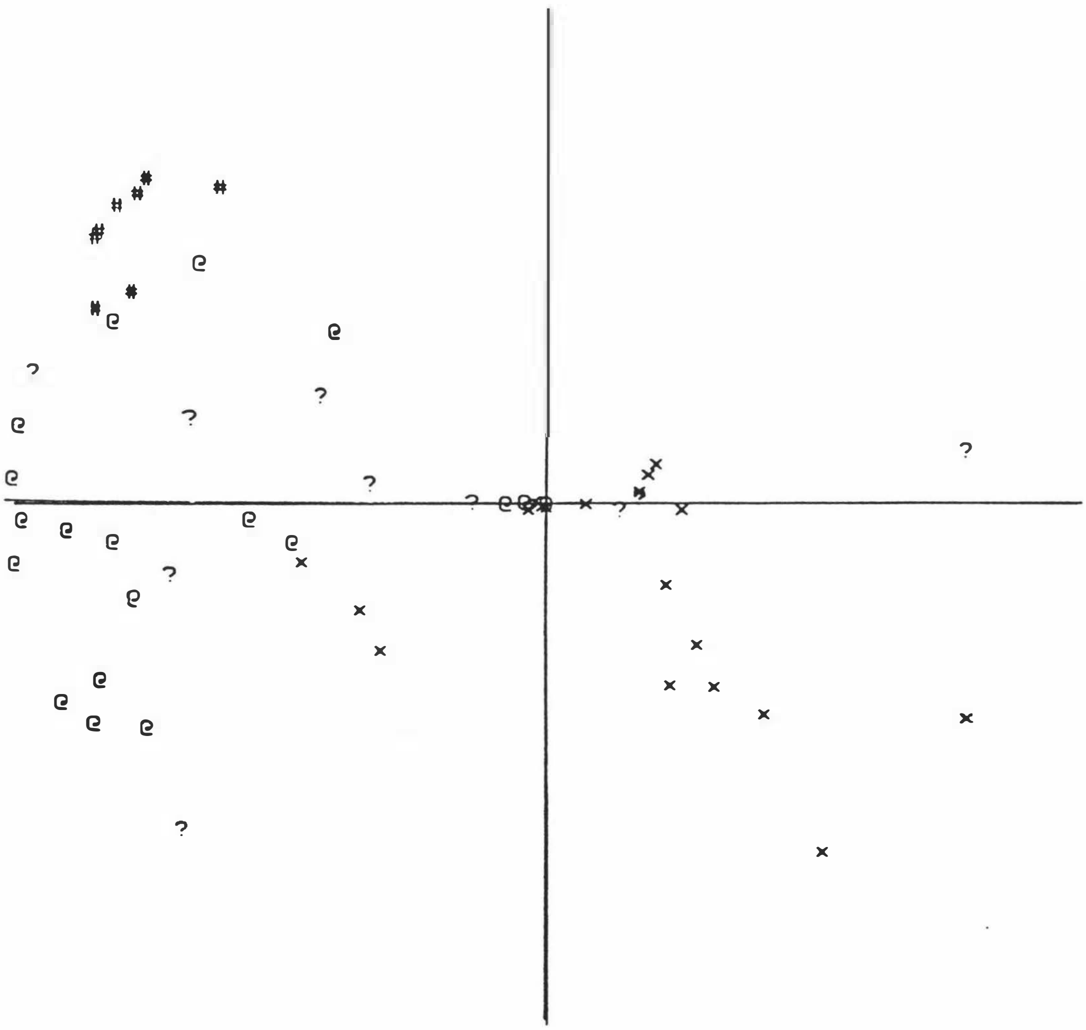


Figure 4.9 Subject space for two sets of D15 stimuli, two kinds of data, IDIOSCAL model.

Key: @ Protanopes # Deutanope x Normal vision or undiagnosed.
Data sets from D15 procedure on overlay. Data sets from hierarchical sorting on lower sheet.

The Points-of-View model

I should also mention the Perspectives model [Lingoes & Borg, 1978]. This provides for a different form of individual variation: the private spaces differ in how far each stimulus lies from the centre of the configuration, though the *direction* of the stimulus is fixed (so there are an additional N degrees of freedom per subject). Lingoes and Borg incorporated this and previously-described models in a framework, PINDIS, where they are arranged in increasing generality (requiring progressively more degrees of freedom).

The options presented by all these models for distorting a group configuration into private configurations are still restricted. Indeed, this is part of their appeal. The assumptions they make are precise enough to be falsifiable. The problem is, deciding *a priori* which model to apply. For colour vision, reasons for using the IDIOSCAL method were known before analysing the data, but this is seldom the case.

If the data are abundant, one can scale each data set in isolation, producing M solutions, then scale the M-by-M matrix of congruence coefficients between those solutions (in other words, scaling a matrix of distances between points in configuration space[3]). Often the arrangement of points can be approximated in far fewer than (M-1) dimensions. This was the approach taken by Kirk and Burton [1977]. In that example, their informants responded to triads data for 13 personality descriptors, in $M = 8$ situations: the descriptors were presented in isolation, in one test, while the other seven tests consisted of applying the same descriptors to particular 'social identities' in Maasai culture. Kirk and Burton compared the separate solutions (semantic structures) organised around these (7+1) social identities, and found that a two-dimensional map was adequate for portraying the dissimilarities between them.

One might discover, in fact, that the arrangement of the solution points in configuration space is more-or-less one-dimensional. In such a situation, the picture can be simplified by fitting a regression line through the points, and replacing each individual solution with the point on the line closest to it, thereby fitting the data with only $M + 2\,N\,P$ degrees of freedom instead of

[3] Borg and Leutner [1985] explain why the congruence coefficient, instead of the product-moment correlation, should be used to assess the similarity between two configurations.

$M N P$. This is a form of the Points-of-View model, introduced by Tucker and Messick [1963]: an especially general alternative to the preconceptions of the INDSCAL model.

In this account the "viewpoints" are the extremes of variation, alternative perceptual maps in their purest form; these are mixed, in various linear combinations, to model the subjects' responses. In the simple situation I'm concentrating on, there are two viewpoints, X' and X'', defining a spectrum of configurations,

$$X_m = a_m X' + (1 - a_m) X'' \qquad (0 \leq a_m \leq 1) \qquad (4.9)$$

Points on that line correspond to *idealised individuals*, more or less compatible with the observations of actual subjects. Idealised individuals should not be confused with viewpoints.

Another name for viewpoints, "sources of dissimilarity" [Meulman & Verboon, 1993] is perhaps misleading, since it focusses attention on the inter-stimulus distances. Ross [1966] emphasised that mixing the *distances* between items in two viewpoints is not valid; one must take linear combinations of the *coordinates*. Not just any vector of $N (N\text{-}1) / 2$ numbers can be distances between N points – there are very tight constraints to be satisfied. Though two vectors may satisfy them, in general a linear combination of the vectors will not.

Modelling the data as accurately as possible becomes a matter of optimising the whole spectrum by altering X' and X'', and the subjects' mixture parameters a_m, concurrently. The way this is done in the Tucker-Messick analysis is equivalent to performing a matrix decomposition on the data. The data are assumed to be at ratio level, and the analysis is a metric one. I hope to convince the reader that like the INDSCAL model, the Points-of-View model can be applied to triadic data (and other forms of distance comparison) directly.

The general case of V points-of-view confines idealised-individual solutions to a (V-1)-dimensional subspace of the configuration space, defined by a simplex having the points-of-view for apices. $V = 1$ returns us to the simplest case of a single configuration. This reduction in dimensionality is the key point. It allows scaling to proceed, even when the data from each subject are too sparse to process in isolation: constraints are placed on X_m by the equally-sparse data from the other (M-1) subjects, indirectly, with the viewpoints as intermediaries.

Meulman and Verboon [1993] present an interesting variation of this model, in which subjects are assumed to embody one or other of V points-of-view; there are no intermediate mixtures. This comes back to assuming that a subject's data arises from a randomly-perturbed replication of a shared configuration, but allowing more than one such configuration. So the points-of-view partition the subject group into V non-overlapping, internally homogeneous sub-populations. To further simplify the model, the V configurations are assumed to differ only in dimensional saliences. This variation has much in common with the Latent Class model [Winsberg & de Soete, 1993]. It has the advantage that points-of-view are uniquely specified, whereas in the Tucker-Messick formulation, points of view are not unique. For instance, setting $V = 2$, constraining individual solutions to line on a line segment through configuration space, does not specify the segment's end points X' and X''; there are infinitely many pairs to choose from. Multiple solutions exist to the matrix decomposition.

However, the emphasis in Tucker and Messick's paper is on continuous variation, rather than distinct populations. This seems to me to be closer to the spirit of developmental psychology. I return now to that original form of the model [used in a number of early studies: Helm & Tucker, 1962; Landis, Silver, Jones & Messick, 1967; Silver, Landis & Messick, 1966].

Note, first, that end-points are not necessary. X' and X'' have the purpose of delimiting a line segment, but it is equally easy to define one point (the group configuration: call it X_0), and a line passing through it in the direction Y, so that $X = X_0 + a\ Y$. Thus the configuration for subject m (the m-th idealised individual) is

$$X_m = X_0 + a_m\ Y \tag{4.10}$$

This is still the points-of-view model, but without the points of view. From (4.10),

$$d_{m,ij}^{\ 2} = \sum_p (x_{m,ip} - x_{m,jp})^2 \qquad = \qquad \sum_p((x_{0,ip} - x_{0,jp}) + a_m\ (y_{ip} - y_{jp}))^2$$

i.e. $$(\mathbf{x}_{m,i} - \mathbf{x}_{m,j})^{\mathrm{T}} \cdot (\mathbf{x}_{m,i} - \mathbf{x}_{m,j}) \qquad = ((\mathbf{x}_{0,i} - \mathbf{x}_{0,j}) + a_m\ (\mathbf{y}_i - \mathbf{y}_j))^{\mathrm{T}} \cdot ((\mathbf{x}_{0,i} - \mathbf{x}_{0,j}) + a_m\ (\mathbf{y}_i - \mathbf{y}_j))$$

Inserting this into (2.9) gives the derivatives of S_m, the m-th contribution to raw Stress, as

$$\partial S_m / \partial \mathbf{x}_{0,i} \qquad = \sum_j f_{m,ij}\ (\mathbf{x}_{m,i} - \mathbf{x}_{m,j}) \,/\, d_{m,ij} \tag{4.11a}$$

$$\text{where } f_{m,ij} = 2 \sum_{(k,l)} \varepsilon_{m,ij,kl}\ \Theta(d_{m,kl} - d_{m,ij})$$

$$\partial S_m / \partial \mathbf{y}_i \qquad = a_m \sum_j f_{m,ij} (\mathbf{x}_{m,i} - \mathbf{x}_{m,j}) / d_{m,ij} \qquad (4.11b)$$

$$\partial S_m / \partial a_m \qquad = a_m \sum_{(i,j)} f_{m,ij} (\mathbf{y}_i - \mathbf{y}_j)^{\mathrm{T}} \cdot (\mathbf{x}_{m,i} - \mathbf{x}_{m,j}) / d_{m,ij} \qquad (4.11c)$$

In the more general case of $V > 2$, where $(V+1)$ dimensions are required to represent how the individual solutions are arranged in configuration space, one could write $X = a_m Y + b_m Z + \ldots$ This can be expressed more conveniently as a matrix multiplication: $\mathbf{X}_m = \mathbf{a}_m \cdot \mathbf{Y}$, where $\mathbf{a}_m$ is a row-vector of $(V\text{-}1)$ elements, while the columns of $\mathbf{Y}$ are Y, Z, etc.

The INDSCAL model in P dimensions is simply one way of confining individual solutions to a $(P\text{-}1)$-dimensional subspace of configuration space. The principal difference between it and the Points-of-View model is that the orientation of that subspace (a line or a plane, for $P = 2$ or 3) is established in advance, instead of being determined by the data. Other models offer subspaces with different alignments.

An interesting special case of the Points-of-View model is the simple situation of $V = 2, P \leq 2$. Consider X_m, the configuration for the m-th idealised subject: N points arranged in P dimensions. I find it helpful to imagine this arrangement as a horizontal cross-section through a $(P+1)$-dimensional space, where the vertical axis, the $(P+1)$-th, represents a. The lines passing obliquely through figure 4.10, a hypothetical example for $P = 1$, are the possible positions for the N stimuli; their intersections with a horizontal plane at a_m are the elements' positions in X_m. MDS on data from multiple subjects is a matter of arranging the lines and the heights of the cross-sections so as to best account for the relationships observed within each section.

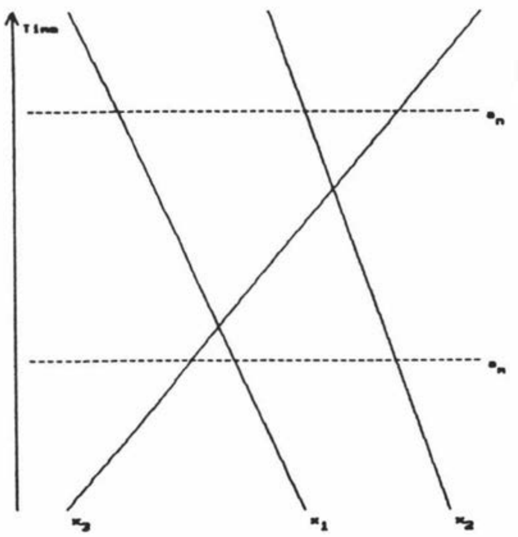


Figure 4.10 Points-of-view model with $V = 2$, $P = 1$

Some examples which can be visualised in this way come from studies of social networks (a college fraternity, in Nordlie [1958]; a monastery, in Sampson [1968]), in which the personal distances between persons (the elements) evolved in the course of repeated observations. The vertical axis in such cases would be a time dimension, with each a_m slice being a 'snapshot' of a developmental process.

The mathematics is the same for situations in which it is the stimuli or the relationships between them which change, rather than the subjects assessing them.

In the Points-of-View model, there is no connection between the slope of one oblique line (stimulus) and another – they pass through the $(P+1)$-dimensional space independently of each other. This is another way of clarifying the restrictions built into models such as IDIOSCAL, where the lines are coordinated and coherent.

Just as with INDSCAL, this model retains ambiguities of scale, which need to be squeezed out in the course of minimising S. X_0 is periodically renormalised to keep Scale(X_0) constant.

Although I have glibly described the X_m as single points in configuration space, an ambiguity remains, in that rotated versions are equally good solutions, not to mention reflections and dilations. It is possible, but undesirable, for Y to have a rotational moment. Reverting for a moment to the idea of the range of allowable X being defined by view-points X' and X'', consider the case where X' is a copy of X'', rotated so that $x'_{ip} = -x''_{ip}$. Then configurations mixing these viewpoints vary in scale – X_m in (4.9) collapses to a single point when $a_m = 0.5$.

Rotational and dilational components of Y can be eliminated in the course of the scaling. In the $(P+1)$-dimensional model, the stimulus lines should on average be vertical, with no net rotational moment. To put it another way: the transformation mapping X_0 onto X_m should be Procrustean.

There are three other constraints to be imposed on Y in order to remove forms of non-uniqueness. First, let $\Sigma\, y_{ip}^2 = 1$, i.e. $\|Y\| = 1$. Second, let $\Sigma a_m = 0$, i.e. X_0 should be 'central'. Third, in the case of $V > 2$, the columns of $\mathbf{Y}$ should be orthogonal.

For the data sets presented so far, this form of analysis leads to the same results as the INDSCAL model. The same is true for triadic data which will be presented in Chapter 8, where the stimuli were tape-recordings of canine heartbeats, compared for dissimilarity by various members of the Massey veterinary faculty.

A Residual-Forces Subject Scaling

This thesis concentrates on sparse data sets, for which MDS solutions in isolation cannot always be derived. This rules out the first stage of individual-variations analysis, that of scaling the differences between solutions. Another tactic which is sometimes used – scaling differences calculated directly between the responses in each data set – is seldom applicable either, since the data will often have been elicited to minimise the overlap between data sets so that they complement one another.

I suggest that even with data sets too sparse to be usefully compared with each other, it is still possible to measure the extent of each set's disagreement with the solution derived by scaling them in combination, and more importantly, to measure the *direction* of that disagreement.

When the downhill descent reaches a minimum of Stress, the m-th data set will still conflict with X (unless $S_m = 0$). It contributes ∇X_m, pushing X in a direction which will bring it more into agreement with that subject's response ($\Sigma\ \nabla X_m = 0$ since the downhill descent has converged). Similar data sets will disagree with the group configuration in a similar way, and exert roughly parallel forces. It turns out that in many cases, the ∇X_m (vectors in configuration space) can be scaled in lieu of individual solutions, taking the Euclidean distance between ∇X_m and ∇X_n as a first approximation to the dissimilarity between the m-th and n-th data sets. As well as providing a starting value of $Y^{(0)}$ for the Points-of-View analysis, this is interesting in its own right.

Applying this 'residual forces' approach to the Shepard-interval sound experiment of Chapter 3, where the data are not complete and each subject ($M = 9$) has responded to a different list of triads, the residual forces at equilibrium can be compressed into two dimensions, with the dominant form of individual difference coinciding with variations in dimensional salience revealed by INDSCAL: figure 4.11(a). In the case of the Stalmeier-de Weert 16-colour triad data, where $M = 10$, I found that a two-dimensional solution again was sufficient (figure 4.11(b)). Most of the differences between the ∇X_m are along the first dimension, which coincides with the variations in dimensional salience. The second dimension reflects curvatures in the subjects' private colour spaces which cannot be obtained by stretching or compressing a group space (figure 4.12).

Figure 4.11 Subject spaces for Stalmeier-de Weert data (a, left) and Shepard-interval triads (b, right), mapped using residual forces.

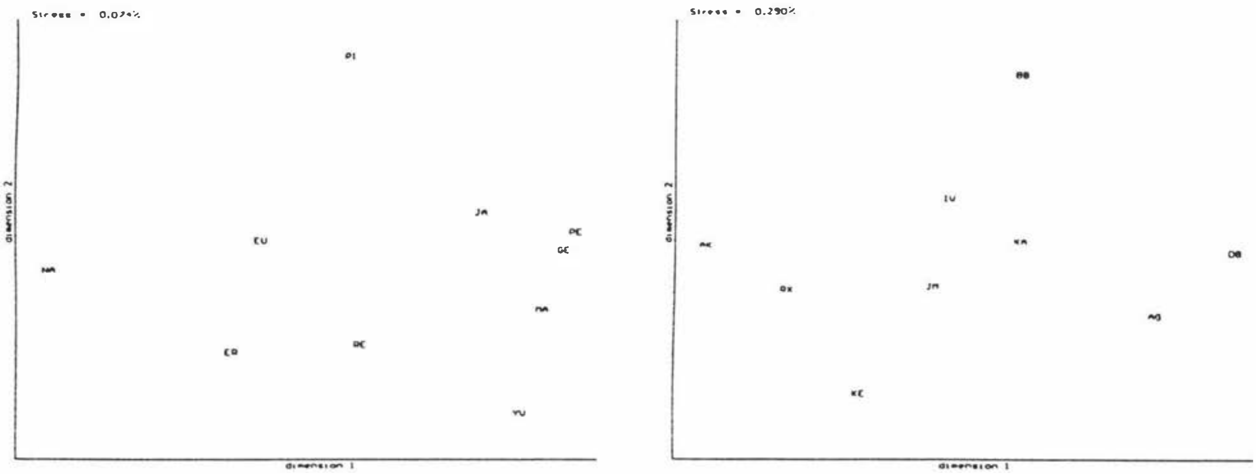
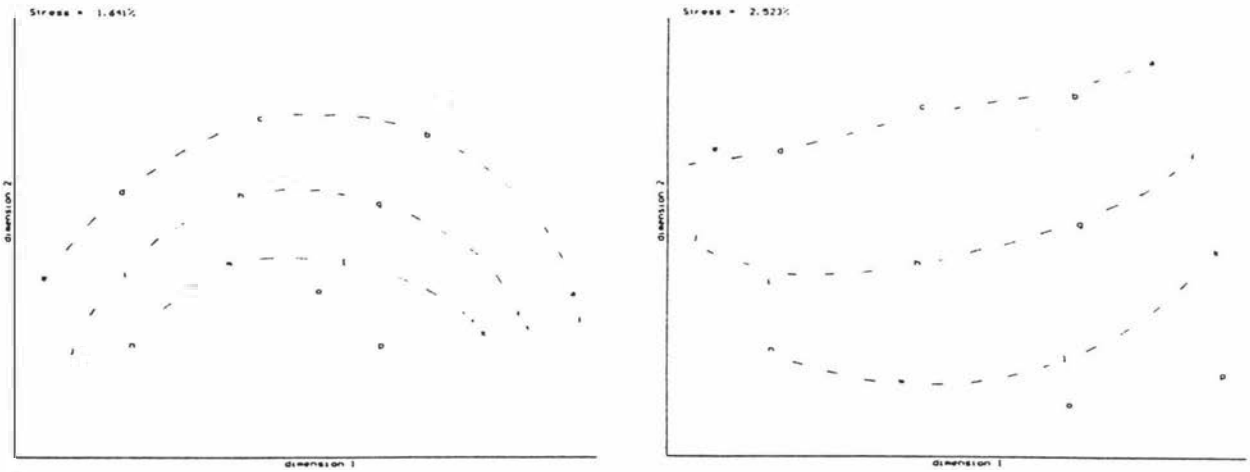


Figure 4.12 Individual solutions for Stalmeier's subjects PI and YV, illustrating forms of individual variations not covered by dimensional weighting



Divers Measures

It is not at all obvious that perceptual spaces should all be governed by the Euclidean metric. Attneave [1950] was the first to raise doubts. The only intrinsic requirement for a function to be a "distance" is that it satisfy the triangle inequality, $d_{ij} + d_{jk} \geq d_{ik}$. A small, tentative step in the direction of generality, avoiding introducing too many parameters, brings us to a family of distance functions known as the Minkowski r-metrics. In general, in a P-dimensional Minkowski r-metric, the distance between points i and j is

$$d^{(r)}_{ij} = [\sum_{p=1}^{P} |x_{ip} - x_{jp}|^r]^{1/r} \qquad (4.12)$$

(which is not easily expressed in matrix form). For $r = 2$ this becomes the standard Pythagorean equation for distance.

MDS theory interprets these r-metrics as a family of alternative rules for combining differences along underlying dimensions to arrive at the perceptual dissimilarity.

At one stage, researchers hoped that the appropriate combination rule for the dimensions of a given perceptual domain could be determined by scaling dissimilarity data, inserting a variety of distance definitions ("metrics") into the Stress equation and seeing which one resulted in the lowest Stress. However, it is known now that Stress values are not comparable across metrics [Shepard, 1974]. The question must be settled in other ways (Shepard and Cermak [1973] plotted the isosimilarity contours directly, in a known configuration of complex figures, to reveal a roughly city-block metric). The choice of metric for scaling data will generally be determined *a priori*, reflecting one's knowledge (or assumptions) about the underlying psychophysics, more than it sheds light thereon.

Since Attneave, it has been traditional to distinguish between analyzable (*separable*) and non-analyzable (*unitary*, integral) stimuli[4]. In the unitary case, one assumes that the dimensions of

[4] The nomenclature for these non-Euclidean cases is muddied by the failure of early researchers to consider the cases where $r > 2$. Garner and Felfoldy [1970] only consider two cases, the Euclidean and the city-block metrics, labelling the former "integral" and the latter "non-integral", basing the distinction on the independence of dimensions in the latter (i.e. lack of interference: the absence of facilitatory or inhibitory effects during a constrained classification task), versus the dependence observed in the former. This leaves open the question of describing intermediate cases, or $r > 2$, where there is certainly interference between dimensions.

the perceptual space are not considered separately in the process of assessing dissimilarities. The definitions are couched in words such as "holistic". An example of unitary stimuli is locations within familiar physical space, where the distance from a chair to a particular corner of the room is more likely to be measured directly than by combining two displacements, each measured parallel to one side of the room.

Even if one were to separate dimensions, measuring along the axes of the room, any other orthogonal set of axes would serve equally well. So $r = 2$ seems appropriate in this case.

Conversely, metrics with $r \neq 2$ apply to analyzable stimuli. These have qualities which can be considered in isolation: for example, duration, loudness, and frequency of a musical note. It may be that the recipients of the stimuli are representing these qualities separately, with verbal or conceptual labels. The important point is that the dimensions of the perceptual space – the possible ways one stimulus can differ from another – are distinguishable, their contributions to the dissimilarity easier to disentangle.

In one limiting case of dimensional independence, the contribution to the dissimilarity from a displacement $(x_{ip} - x_{jp})$ on the p-th dimension is the same, whatever the displacements on other dimensions. Then the dissimilarity is the sum of absolute displacements:

$$d^{(1)}{}_{ij} = \sum_{p=1}^{P} |x_{ip} - x_{jp}| \,, \tag{4.13}$$

the r-metric with $r = 1$, or *city-block* metric.

At the other extreme of unitary stimuli is $r = \infty$, the supremum or dominance metric:

$$d^{(\infty)}{}_{ij} = \max_{p=1}^{P} |x_{ip} - x_{jp}| \tag{4.14}$$

Here, the contribution from the p-th dispacement is either 0 or $|x_{ip} - x_{jp}|$, depending on whether or not the (absolute) displacement on some other dimension p' is larger. The greater r, the greater the degree to which a dyad's dissimilarity is dominated by the dimension in which the greatest displacement occurs (this dimension varying from dyad to dyad).

Oddly enough, in two dimensions these two limiting metrics are equivalent. Consider two configurations, identical apart from a 45° rotation. We calculate the same distances between elements (apart from a constant factor) when we apply the city-block metric to either configuration and the supremum metric to the other.

This illustrates a key aspect about the non-Euclidean r-metrics. As with the weighted Euclidean metric, rotational invariance of distance is broken, and consequently, some choices of axes result in lower values of Stress than others.

In practice, the distinction between unitary and analyzable stimuli is not so clear-cut. Unitary dimensions can be made analyzable, in tests designed to focus attention on their distinctive qualities [Burns & Shepp, 1988; Stalmeier & de Weert, 1991b]. Even when a subject does not report being aware of the dimensions in isolation (for instance, hue and saturation of areas of colour), he or she may still be processing them separately at a level below awareness. Stimuli such as the consonants of speech, with clearly analyzable characteristics, become unitary in confusion-matrix experiments [Miller & Nicely, 1955] where subjects are prevented from consciously attending to the auditory features. A quasi-sorting task applied to the same consonants [Pruzansky, 1969], in which subjects perhaps separate the dimensions to label or conceptualise the stimuli, resulted in a different configuration, as have studies where subjects assessed the similarity between consonants directly.

To add to the confusion, hybrid situations are conceivable, with some of the dimensions unitary, while others are analyzable. A simple example would be physical space, with altitude as a qualitatively different dimension from the North-South and East-West axes, unlikely to be mistaken for either. Such metrics may sound artificial and far-removed from psychophysics, but mentioning them is not a case of obfuscation for its own sake, since a similar distinction between classes of dimensions has been proposed for pain experiences [Torgerson & BenDebba, 1983]. In the case of colour space, brightness is qualitatively different from the hue and saturation dimensions. See also the stimulus-specific dimensions proposed by Winsberg and Carroll [1989].

Shepard [1964] contrived stimuli where the metric is not a constant function, but varies from point to point in the perceptual space, superimposed on individual variations (dimensional weights). In this case, it is the values of stimuli on a dimension, rather than the experimental design, which focus attention on displacements along that dimension. Shepard remarked that for all their apparent generality, the family of r-metrics are little less specialised than the Euclidean case.

Implementation

It is a relief to return from these contentious areas to the definition of Stress, and to find the arithmetic of non-Euclidean MDS to be far simpler than the arguments as to its applicability. I derive them, to show that the metrics are compatible with the pairwise-comparison definition of Stress as well as the Kruskal form, not because of any vociferous demand for the options.

The implications of altering the definition of "distance" are two-fold. First, by (4.12), the Stress (i.e. in the spring model, the potential energy stored by compressing and stretching springs) is changed: some distance relationships which previously agreed with the subject's judgements do so no longer, and *vice versa*. Secondly, the direction in which the (i,j)-th spring exerts its corrective force on i and j is different as well, which the calculation must take into account while in its second phase of iterating over (i,j) to accumulate the total force on the i-th item.

Springs push and pull radially, but the word "radial" needs re-defining in non-Euclidean space where "circles" (the set of points equidistant from a central point) are non-circular. Shepard terms them *isosimilarity contours*. The only requirement for an isosimilarity contour is that it be convex, in order to satisfy the triangle inequality. In the supremum metric, a "circle" of radius r is a square. In the city-block metric it is a rhombus (reminding us of the equivalence of the two extremes. This equivalence fails in two or more dimensions, with an isosimilarity surface becoming cubic in the supremum metric and octahedral in the city-block metric).

Re-phrase spring behaviour thus: the force on element i from the (i,j)-th spring force $f^{(r)}_{ij}$ is exerted perpendicularly to the surface of a hypersphere passing through i and centred on j. Force components are diagonal in the city-block, and axial in the supremum metrics.

$f^{(r)}_{ij}$ resolves into components along the p-th dimension: $f^{(r)}_{ij} = \Sigma\, \hat{e}_p f^{(r)}_{ij,p}$

$$f^{(r)}_{ij,p} = r f^{(r)}_{ij} \; \mathrm{sgn}(x_{ip} - x_{jp})\, |x_{ip} - x_{jp}|^{r-1} \, [\Sigma\, |x_{ip} - x_{jp}|^r]^{1/r-1} \qquad (4.15)$$

Obtaining the gradient of S gives the same results:

$$\partial S(X) / \partial x_{ip} = \sum_j (\partial S(X)/\partial d_{ij}) (\partial d_{ij}/\partial x_{ip})$$

$$= \sum_j \sum_{(k,l)} \varepsilon_{ij,kl} \Theta(d^{(r)}_{kl} - d^{(r)}_{ij}) \ \mathrm{sgn}(x_{ip} - x_{jp}) |x_{ip} - x_{jp}|^{r-1} [\Sigma |x_{ip} - x_{jp}|^r]^{1/r-1}$$

which reduces to particularly simple forms at the limiting metrics:

$$r = 1, \qquad \partial S(X) / \partial x_{ip} = \sum_j \mathrm{sgn}(x_{ip} - x_{jp}) \sum_{(k,l)} \varepsilon_{ij,kl} \Theta(d^{(1)}_{kl} - d^{(1)}_{ij}) \qquad (4.16)$$

i.e. dimensions contribute to Stress independently. There are convergence difficulties in minimising this Stress [Arabie, 1991]: the optimal order of points must be sought along each dimension in isolation, because of their independence. In effect there are P separate one-dimensional configurations to minimise. Points cannot swerve out of one another's way in the course of the convergence, and are thus more liable to become trapped in local minima. Combinatorial algorithms are more appropriate than gradient ones.

When $r = \infty$,
$$\partial S(X) / \partial x_{ip} \begin{cases} = \mathrm{sgn}(x_{ip} - x_{jp}) \sum \varepsilon_{ij,kl} \Theta(d^{(\infty)}_{kl} - d^{(\infty)}_{ij}) & \text{if } p \text{ maximises } |x_{ip} - x_{jp}| \\ = 0 \text{ for other } p. \end{cases} \qquad (4.17)$$

(only a single dimension in each dyad contributes to Stress).

Non-circular isosimilarity contours have already appeared in the weighted Euclidean model, where they are elliptical. In the IDIOSCAL model they are ellipses with subject-specific orientations. Thus (4.1) and (4.7) conceal another example of peculiarly-behaved spring force directions. In passing, I note that within the weighted Euclidean model, a group of subjects can be contrived whose weights are distributed so that their isosimilarity contours average to mimic a $r < 2$ metric [Fischer & Micho, 1972]. A situation seemingly governed by the city-block metric may turn out, upon applying the INDSCAL model, to be Euclidean. More generally, when individual weights vary, and dissimilarity matrices Δ_m are averaged, the effect is to decrease the apparent r. "[F]rom an averaged metric little can be inferred about individual metrics, attention distributions, mental mechanisms" [Fischer & Micko].

Examples of $r < 2$ are common though sometimes contrived [Hyman & Well, 1968]. When the dimensions are completely separable, in the limiting city-block case, they can be studied in isolation, and there seems little need for MDS. For examples of $r > 2$, see Gregson's studies of taste [1966a, 1966b] and Arnold's monograph on semantic distances [1971].

There is no barrier to combining non-Euclidean metrics with dimensional weights (when "circles" would be rectangular or rhomboidal), but nor is there any compelling need to do so.

There is a final reason for caution in the use of the two limiting-case metrics, worth mentioning, to back the earlier claim that Stress minima are not comparable across metrics. This is a geometrical feature [Shepard, 1974]. We have seen the isonormal surfaces to be (hyper)-cubic in the Supremum case – with the result that 2^P elements can be placed on the corners of a P-dimensional hypercube, all equally distant from each other (as opposed to P+1 in the Euclidean case). The Stress for such a configuration is zero, whatever the data. It is a trivial configuration, containing no structural information. Similarly, $2P$ elements at the corners of a P-dimensional hyper-octahedron are equidistant in the city-block metric.

The next chapter examines the problem of degenerate solutions in more detail.

5. MAXIMUM LIKELIHOOD ESTIMATION

Maximum Likelihood Estimation is a long-established statistical approach to problems which involve parameters, to be estimated, underlying data which are not complete. Ramsay [1977, 1978] championed its application to the specific problems of MDS, where the parameters are the elements' coordinates in perceptual space. For Ramsay's program, MULTISCALE, the data are dissimilarities. For a version of MLE handling partially rank-ordered directional data (conditional rank-ordering; also triads), see Takane and Carroll [1981].

The implementation of MLE discussed in this chapter preserves the conceptual framework I have already constructed: the extended Johnson pairwise approach to MDS, interpreted by the spring model. Ramsay's presentation follows a different track, though the destination is the same.

I will not delve too deeply into the theoretical foundations. The central concept, a particular application of Bayesian probability, is taken as axiomatic. In normal uses of probability, one would calculate the probability of observing each of an array of potential data sets, when a particular configuration is given. But in MLE the data set is given; in a reversal of causality, one derives the most likely configuration, X_0, such that

$$\Pr(X_0 \mid \text{data}) > \Pr(X' \mid \text{data}) \text{ for all other } X'.$$

According to Bayesian reasoning, we can search configuration space, and find the configuration Y_0 for which the probability of producing the observed data is maximal,

$$\Pr(\text{data} \mid Y_0) > \Pr(\text{data} \mid Y') \text{ for all } Y' \neq \text{Y}$$

then Y_0 is the best available estimate of X_0. This is not obvious.

Before proceeding with the details of the search, the reader might ask, is there a need? What advantages can MLE offer over the Stress-minimising algorithms considered so far? What are the disadvantages of the latter?

One useful feature of the Maximum Likelihood approach is the ability to draw a confidence region in configuration space around some X and say, there is a 95% chance of the true answer lying somewhere therein. Of course the region is more conveniently portrayed as P-dimensional ellipses around the N items in the perceptual-space representation of X, as offered by MULTISCALE. The size of these confidence ellipses is chastening to those of us who have ever built an elaborate theoretical superstructure upon the exact positions of the stimuli in an MDS output.

The way to highlight the deficiencies of Stress-based MDS is by contriving pathological situations. Imagine three stimuli, A, B and C. Information is limited to the following: 19 observers report that B is more similar to A than to C, $(A,B) \ll (B,C)$, while a single dissenter considers that B and C are the more similar pair, i.e. $(A,B) \gg (B,C)$. In the absence of other constraints on the configuration, Stress is minimised by arranging the points at equal intervals. This offends common sense. Common sense tells us that $d_{AB} < d_{BC}$, how much smaller depending on the "spread" in the observers' judgments. Torgerson's triad analysis agrees. In this regard, the advent of Stress-minimising non-metric MDS was a retrograde step.

The assumption of data sets from multiple subjects is not an essential part of the argument. The whole Stress-minimising approach to MDS for a single subject works by balancing dissimilarity comparisons made explicitly against comparisons implied by the totality of judgments. Perhaps the subject judges that $(A,B) \ll (B,C)$, while the relationship $\delta_{AB} > \delta_{BC}$ can be inferred from other comparisons when they are embedded in Euclidean space. Without such inferred judgments, scaling would be stopped by incomplete data, such as the absence of a comparison between (A,B) and (B,C).

Whether the conflicting comparisons are explicit or inferred, the algorithms so far described have a tendency to reconcile the conflict by assigning the same value to all the distances involved. Stress is thereby minimised, in a trivial way that yields no useful structural information as to the true arrangement of stimuli. Such solutions are degenerate. Sometimes their uninformative nature is obvious, as when the distance assigned to a number of dyads is zero, as the elements collapse to a single point; but it is just as unhelpful if they arrange themselves in an equilateral triangle [Borg & Lingoes, 1987; Shepard, 1974]. Figure 4.1(b) is an example of such degeneracy. The remedies put forward by Borg and Lingoes owe more to pragmatism than to theoretical rigor.

In its early optimistic years, non-metric MDS was seen as potentially a method of "purifying" data; it was hoped that the configuration derived from fallible, noisy data would be a more reliable way of predicting the subjects' future judgments on given comparisons than previous judgments, from the same subjects, considered in isolation. Isaac [1970] conducted ingenious "odd-one-out" experiments to test this prospect, within a perceptual space of facial expressions, with negative results. Alas, the more fallible the subjects, and the greater the need for data purification, the greater the chance of a degenerate (and unhelpful) solution arising from the reconciliation of the conflicting comparisons.

I return now to the question of improving the way conflicts are reconciled. The reasoning has much in common with early, metric approaches to MDS: Torgerson's treatment of triads, and Messick's analysis of Successive Intervals data.

Two crucial assumptions about the comparisons between dyads will be made, to bring them under the rubric of the Law of Comparative Judgment (Thurstone's Case V). This predicts that the probability of the informant judging (i,j) to be a less similar pair of stimuli than (k,l) is

$$\Pr\{(i,j) \gg (k,l) \mid D_{ij,kl}\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{x=-\infty}^{D_{ij,kl}} \exp(-x^2 / 2\sigma^2)\, dx \tag{5.1}$$

where $D_{ij,kl}$ is the difference between distances in the perceptual space,

$$D_{ij,kl} = d_{ij} - d_{kl}$$

This is variously known as the normal integral, the error function, or the cumulative density function (c.d.f).

Decreasing fallibility corresponds to $\sigma \rightarrow 0$, with the ideal case ($\sigma = 0$) being a perfectly reliable subject who responds with $(i,j) \gg (k,l)$ (or $\varepsilon_{ij,kl} = 1$) whenever $d_{ij} > d_{kl}$, and $\varepsilon_{ij,kl} = 0$ (while $\varepsilon_{kl,ij} = 1$) otherwise. For $\sigma \neq 0$, $\Pr\{\varepsilon_{ij,kl} = 1\}$ becomes 1 or 0 only asymptopically, for $|D_{ij,kl}| \rightarrow \infty$.

In this model, the probability of the judgment being incorrect (i.e. not in accordance with distances) varies with $D_{ij,kl}$ as well, peaking at 0.5 when $D_{ij,kl} = 0$.

Therefore the likelihood of observing the data set in its entirety,

$$\Pr\{\text{data} \mid X\} = \prod_{(i,j) \gg (k,l)} \Pr\{\varepsilon_{ij,kl} = 1\}$$

When $\varepsilon_{ij,kl} = 0$, either (i,j) and (k,l) were not compared (missing data), or (i,j) « (k,l), in which case there is a factor in the product for $\varepsilon_{kl,ij}$.

It is safe to use log-likelihood instead, since maximising the logarithm is equivalent to maximising the probability itself. For further mathematical simplicity I substitute the logistic function for the c.d.f.

$$\Pr'\{\varepsilon_{ij,kl} > 0\} = \{\exp(-D_{ij,kl}\,\tau) + 1\}^{-1} \tag{5.2}$$

where the parameter τ is approximately $\pi / (3^{1/2}\,\sigma)$.

$$\begin{aligned}
\text{Then } L = \log(\Pr\{\text{data} \mid X\}) &= \sum_{(i,j)} \sum_{(k,l)} \varepsilon_{ij,kl} \log(\Pr'\{\varepsilon_{ij,kl} > 0\}) && (5.3)\\
&= \sum_{(i,j)} \sum_{(k,l)} \mathrm{l}(i,j,k,l) \\
&= \sum_{(i,j)} \mathrm{l}(i,j) \qquad \text{where } \mathrm{l}(i,j) = \sum_{(k,l)} \mathrm{l}(i,j,k,l) \\
\text{and } \mathrm{l}(i,j,k,l) &= -\varepsilon_{ij,kl} \log(\exp(-D_{ij,kl}\,\tau) + 1) && (5.4)
\end{aligned}$$

Clearly $-\infty < L < 0$. As well as being simpler, the logistic function corresponds to Luce's model of choice (choosing which quantity is smaller).

As with Stress, L can be maximised with a gradient algorithm (though following the slope up-hill rather than down).

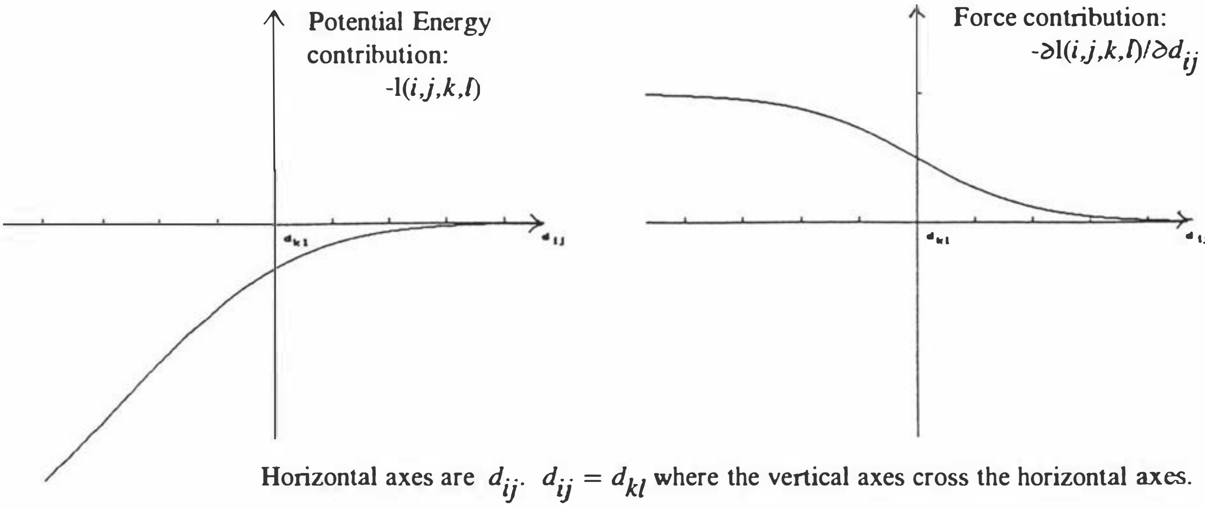
$$\begin{aligned}
\partial L / \partial x_{ip} &= \sum_j (\partial \mathrm{l}(i,j) / \partial d_{ij})\,(\partial d_{ij} / \partial x_{ip}) = \sum_j \sum_{(k,l)} (\partial \mathrm{l}(i,j,k,l) / \partial d_{ij})\,(\partial d_{ij} / \partial x_{ip}) \\
&= \tau \sum_j (x_{ip} - x_{jp}) / d_{ij} \sum_{(k,l)} \varepsilon_{ij,kl}\, \{1 + \exp(\tau\, D_{ij,kl}) + 1\}^{-1} && (5.5)
\end{aligned}$$

Glancing back a few chapters to compare equations (5.4) and (5.5) against (2.5), we find that they are still encompassed by the spring model: the potential energy of the (i,j)-th spring is made up of contributions $-\mathrm{l}(i,j,k,l)$ from comparisons between (i,j) and (k,l); each comparison contributes $-\partial l(i,j,k,l) / \partial d_{ij}$ to the force f_{ij} exerted by that spring on elements i and j. The force contributions are simply the logistic function again (figure 5.1).

$$-\partial l(i,j,k,l) / \partial d_{ij} = -\tau \{1 + \exp(\tau D_{ij,kl}) + 1\}^{-1} \qquad (5.6)$$

$-\partial l(i,j,k,l) / \partial d_{ij}$ asymptopically approaches 0, for $d_{ij} >> d_{kl}$, and -1 for $d_{ij} << d_{kl}$. As τ increases, (5.5) is approximated more and more closely by the step-function force contributions (2.5).

Figure 5.1 Contributions to the (i,j)-th spring's energy (left) and force (right), for logistic equation (τ = 1)



Horizontal axes are d_{ij}. $d_{ij} = d_{kl}$ where the vertical axes cross the horizontal axes.

τ is functioning as a scale parameter. I think of τ^{-1} as the "characteristic length" of errors in the perceptual space under examination: if the absolute difference between d_{ij} and d_{kl}, $|D_{ij,kl}| >> \tau^{-1}$, then the judgment is highly reliable; conversely, if $|D_{ij,kl}| << \tau^{-1}$, the chance of the judgment being erroneous approaches 50%.

There is no denominator to normalise (5.5), so the scale factor is not arbitrary: instead, it is a function of τ. Once a value is assigned to τ, the whole configuration shrinks or expands in the course of maximising L, until τ is in the ratio to configuration scale which best accounts for erroneous judgments (in the earlier sense of judgments conflicting with the reconstruction). τ itself *is* arbitrary. For display purposes, I find it convenient to hold scale(X) constant (by regular renormalising) while allowing τ to expand or shrink. It is just another parameter to optimise over. I use

$$-\partial L / \partial \tau = \sum_{i,j,k,l} \partial l(i,j,k,l) / \partial \tau = \sum_{i,j,k,l} \varepsilon_{ij,kl} D_{ij,kl} \{1 + \exp(\tau D_{ij,kl}) + 1\}^{-1} \qquad (5.7)$$

Figure 5.2 plots this dependence. If all judgments are in accordance with the reconstructed configuration, L can be increased by raising τ (affecting the next iteration of the gradient algorithm, since when the reliability ascribed to each judgment changes, so does the gradient ∇L). Erroneous judgments are indications that τ should be smaller; the weight given to that indication is proportional to the size of the error.

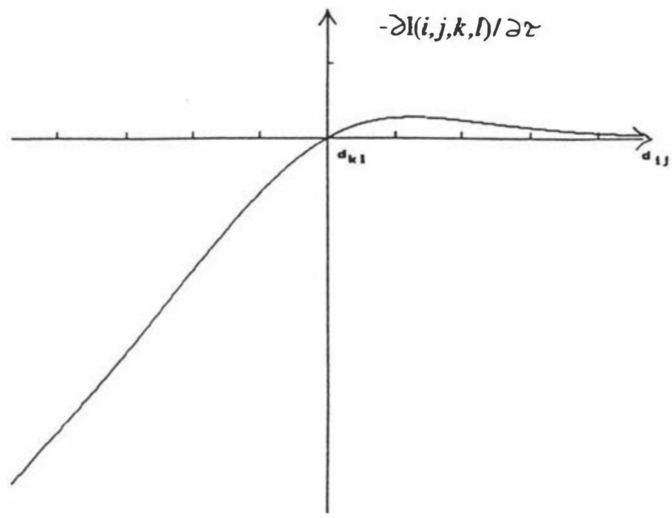


Figure 5.2 Force on τ when (i,j) » (k,l): $\tau = 1$

The only modifications involved in incorporating a Maximum Likelihood Estimation option in MTRIAD are the new spring function, and the optimisation of τ. The models of individual variation covered in the previous chapter, and the non-Euclidean metrics, are not dependent on the shape of the spring function, and retain their usefulness in this new context. I have sought, without success, to find equally simple modifications for the disparity framework used in programs such as KYST. The difficulty is calculating the disparities δ_{ij} themselves, dependent as they are on the particular Stress function being minimised; merely averaging the out-of-rank-order d_{ij} or replacing them with the rank image is no longer enough.

Any of the data sets analysed so far could be re-analysed to illustrate the improvements delivered by MLE. Once again I refer to the POOC study of 13 occupation titles since, as noted in Chapter 3, the data are incomplete. In this example, the configuration is known independently since the POOC files also includes dissimilarity estimates from 286 subjects for 16 occupations which include the 13 scaled with triads. The result of scaling those dissimilarity matrices in combination is figure 5.3. Applying MLE to the triadic data gives figure 5.4, which, compared to figure 3.4(a), is generally more convincing, i.e. closer to 5.3. Note in particular the positions of BSL, MIN and MPN. The value of L for figure 5.3 could be increased by averaging the dissimilarity matrices and concealing the considerable variation between subjects.

Figure 5.3 Dissimilarity estimates for occupational titles, $N = 16$; Maximum Likelihood solution

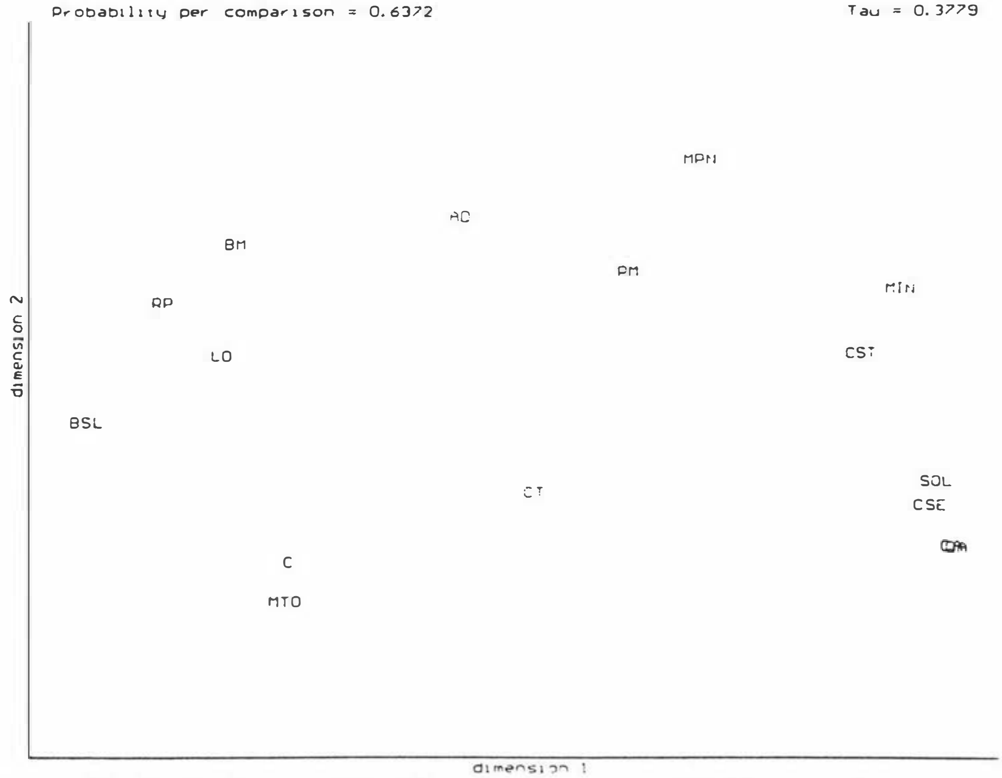
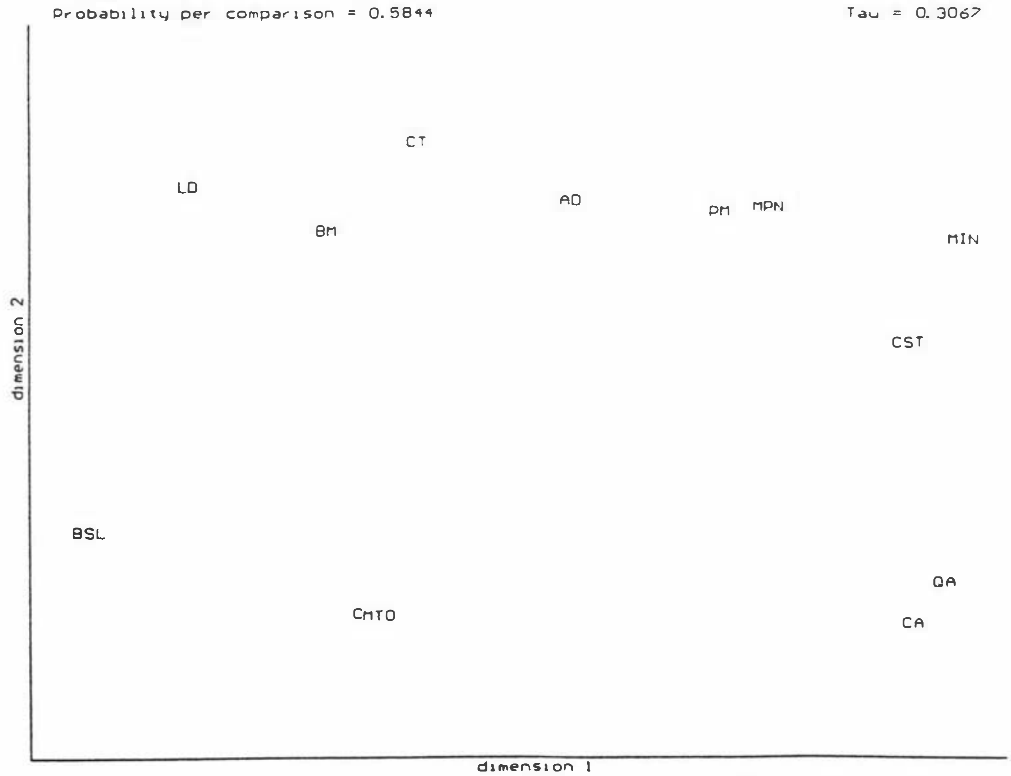


Figure 5.4 Triadic data for occupational titles, $N = 13$; Maximum Likelihood solution



Key to the occupations:

				AD	Ambulance Driver
BM	Barman	BSL	Building-site labourer	C	Carpenter
CA	Chartered accountant	CSE	Civil servant (Executive)	CT	Commercial traveller
CST	Comprehensive school-teacher	LD	Lorry driver	MIN	Church of Scotland Minister
MPN	Male Psychiatric Nurse	MTO	Machine tool operator	PM	Policeman
QA	Qualified Actuary	RP	Railway Porter	SOL	Country solicitor

Included in this figure, in lieu of Stress, are two indicators of the goodness of fit between data and solution. For both, the higher the better. One is τ, in the upper right of the diagram. In the upper left is P, the geometric mean over all comparisons of the likelihoods $l(i,j,k,l)$.

$$P = \exp(L / \Sigma\varepsilon_{ij,kl}) \qquad \text{Clearly } 0 < P < 1.$$

The results of four triad experiments, at the end of Chapter 3, were achieved using the MLE option of MTRIAD. In the fourth experiment the way the data were collected (with an interactive procedure to be described in Chapter 8) left them incomplete as well as unbalanced. Scaling them with Stress rather than MLE leads to barely recognisable results.

A second demonstration uses the Stalmeier-de Weert data. Refer back to figure 4.4(a), depicting the configuration recovered (rather poorly) from 1/8 of the responses from subject PE. The overall trend of the configuration is recognisable, but individual elements have strayed far from their rightful places; the overall crudeness of the recovery reminds us of the guideline for Balanced Incomplete Designs [Burton & Nerlove, 1976], that λ should be at least 2, i.e. that each dyad should participate in at least two triads. Here the design is not balanced but on average $\lambda = 14/8$. Scaling the same data with logistic-function spring-force contributions results in figure 5.5, a closer approximation to the solution for complete data.

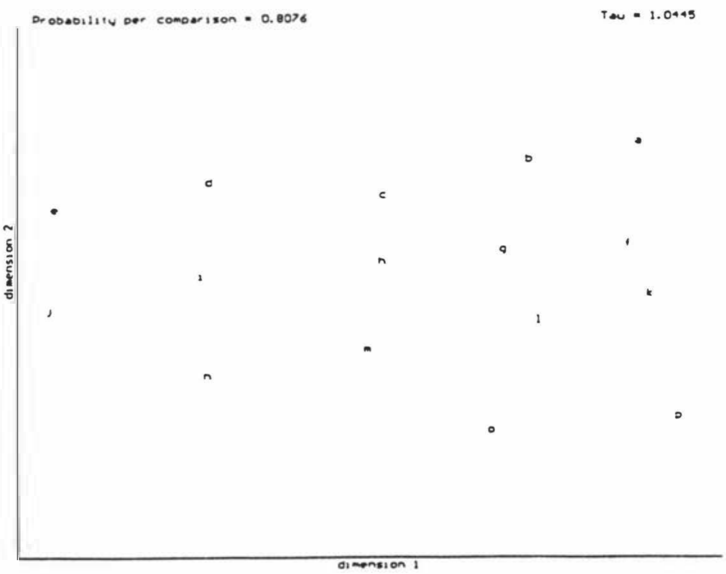
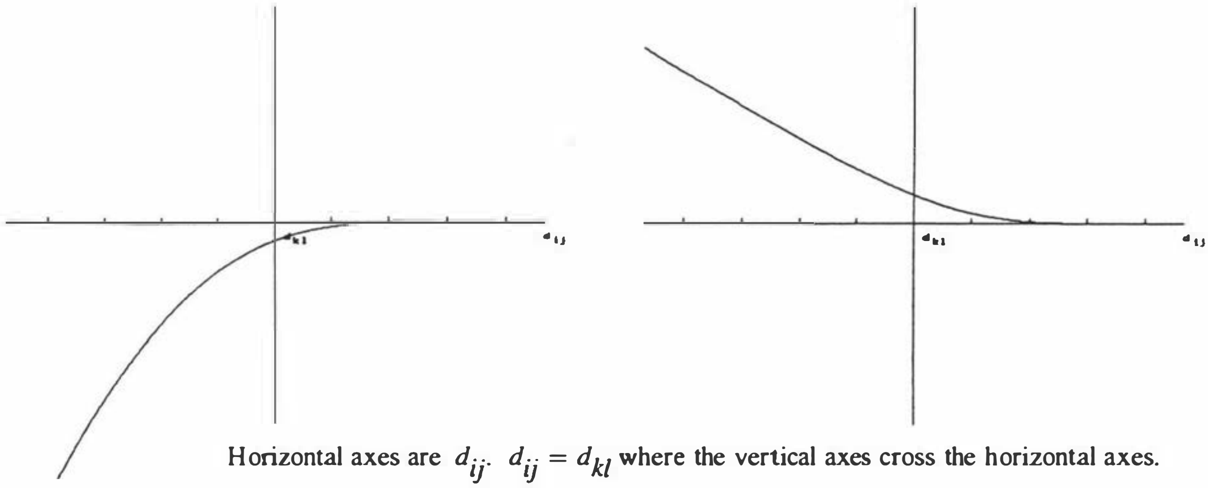


Figure 5.5 MLE solution for incomplete data from PE

Did the earlier simplifying decision to replace the c.d.f. in (5.1) with a logistic function make any difference to the solution?

The log of the normal integral is shown as the solid line in figure 5.6. It is asymptopically flat for $d_{ij} >> d_{kl}$, and asymptopically quadratic for $d_{ij} << d_{kl}$. So the derivative of the curve (dotted line) can be approximated with a ramp function (with slope τ^{-1}).The parameter σ in (5.1) is analogous to τ^{-1} and indicates the accuracy of the approximation. As we approach the ideal case of an infallible subject, σ goes to 0, and the spring-force contributions regain the ramp-function form used to minimise Stress, (2.11). Recall that as τ^{-1} goes to 0, the limiting case of the logistic-force spring force is the step-function spring force. Chapter 2 reported that any differences between configurations reached via these rival definitions of Stress are insignificant. Switching between the two definitions of Likelihood (logistic function and c.d.f.) is equally free from practical effects.

Figure 5.6 Contributions to the (i,j)-th spring energy (left) and force (right), for normal integral ($\sigma = 1$)



Horizontal axes are d_{ij}. $d_{ij} = d_{kl}$ where the vertical axes cross the horizontal axes.

It has long been known, for the comparable situation of unidimensional scaling [Mosteller, 1958], that details of the function G(u) used to model subjects' responses are not crucial. G(u) should be symmetric around 0 (with the corollary that G(0) = 0.5). As u goes to infinity, G(u) should approach 1. Finally, dG(u)/du should be unimodal (by symmetry, the slope is highest at $u = 0$).

The key feature, responsible for the improvement of MLE, seems to be the smooth transition of l(i,j,k,l), as a function of $D_{ij,kl}$, between its asymptopic behaviours at $D_{ij,kl} >> 0$ (where it tapers smoothly to 0) and $D_{ij,kl} << 0$ (a plateau or ramp). For Stress, at $D_{ij,kl} = 0$ there is a discontinuous transition.

The "tail" of the spring force, tapering off smoothly instead of dropping abruptly to zero when $D_{ij,kl} = 0$, is responsible in the pathological ABC situation for shortening d_{AB} at the expense of d_{BC}, even when $d_{AB} < d_{BC}$, the desired effect. Even when they reach an equilibrium, a residuum of spring tension is always present, though in balance.

Note that as τ increases, and the $D_{ij,kl} = 0$ transition becomes sharper (closer to the classical case), the benefit from repeated comparisons (actual or inferred) is less; they have less capacity to push the equilibrium past the $d_{ij} = d_{kl}$ point. When τ is large relative to $d_{ij} - d_{kl}$ for a given i,j,k,l, repetitions of that comparison have low information content because the first observation is reliable. τ might be interpreted as an indication of the redundancy of the data. This has implications for the problem of optimal omission, to be encountered in Chapter 8.[1]

An unexpected benefit of the MLE modification is that it ameliorates the problems of local minima, and of violent oscillations during the iterations, which arise when a gradient algorithm is used to minimise Stress.

Oscillations are a by-product of the simple-minded process by which MTRIAD adjusts the step size in (2.7) to cope with consecutive $\Delta X^{(t)}$. Because of the extreme non-linearity of $S(D_{ij,kl})$, small changes in X can increase $\partial S / \partial X$ dramatically in a feedback loop which the step-size adjustments react too slowly to stabilise.

Local minima were mentioned as a problem in Chapter 2. They are a particular problem when pooling multiple data sets: perhaps because of the increased number of conflicting comparisons. Elements become positioned in a partially degenerate equidistant arrangement so that forces from the conflicting comparisons balance out. In these situations, progress towards accommodating other, non-contradictory judgments require a small, temporary disruption of this equidistance, which is energetically unfavourable (a "potential barrier", in the configuration terrain: see Chapter 2), paralyzing the whole process. Once again, the gradual nature of $l(i,j,k,l)$

[1] One feature of the Lyons *et al* cry data has not been fully exploited yet. In an attempt to find the bounds of subject cooperation, subjects H and I were asked, immediately after making each comparison (primary and secondary), to rate their level of confidence (on a 1-3 scale) in it. This is a rare form of data (another instance of confidence-rated triads is MacRae, Howgate and Geelhoed [1990]), and the question arises of how to use the information. The confidence rating $c_{ij,kl}$ could be taken as an indication of the magnitude of $\Delta_{ij,kl} = \delta_{ij} - \delta_{kl}$, more informative than the "greater than zero / less than zero" binary judgments so far encountered. One can treat a judgment where $c_{ij,kl} = 3$ as equivalent to the same judgment, made three times, and simply weight a comparison's contribution to Stress or Likelihood, by setting $\varepsilon_{ij,kl} = c_{ij,kl}$.

As with explicitly replicated judgments, in the MLE approach this has the effect of lengthening d_{ij} and shortening d_{kl}, even when d_{kl} is less than d_{ij} already. In the Stress minimising approach, the additional comparisons are wasted in the likely event of the configuration agrees with them.

in the vicinity of $D_{ij,kl} = 0$ is the key to solving the problem: small departures from equidistant arrangements are less severely penalised.

It remains to go into detail about the two suppositions required by my rationale for applying the Law of Comparative Judgment.

First, independence of the comparisons is assumed. This assumption breaks down if using the Method of Triadic Comparisons rather than the Complete Method of Triads. However, independence seems to be a good approximation, and I have few qualms about applying MLE to Triadic Comparisons data. Here, there are only six possible rankings of dissimilarities within each triad (the primary comparison within each triad reduces the options for the secondary comparison), compared to eight possible pairwise ways – two of them intransitive – of ranking the dyads when the three comparisons are made independently.

Secondly, I assume that errors in the data conform to the Additive error model [Ramsay, 1977], i.e. that the standard deviation of $D_{ij,kl}$ is a constant (σ in (5.1)), independent of (i,j) and (k,l). The implied model for what happens during each judgment is something like this: the subject some-how assesses distances between elements in his or her mental representation, a process perturbed by a normally-distributed error, and compares the error-perturbed assessments δ_{ij} and δ_{kl}.

$$\delta_{ij} = d_{ij} + e(\sigma / \sqrt{2}), \qquad\qquad \delta_{kl} = d_{kl} + e(\sigma / \sqrt{2}),$$

with δ_{ij} assessed afresh every time (i,j) is involved in a comparison. The observed datum is

$$\begin{array}{ll} (i,j) \gg (k,l) & \text{if } \Delta_{ij,jk} > 0, \\ (i,j) \ll (k,l) & \text{if } \Delta_{ij,jk} < 0, \end{array}$$
$$\text{where } \Delta_{ij,jk} = \delta_{ij} - \delta_{kl} = D_{ij,jk} + e(\sigma).$$

This way of simulating errors has been used in a number of Monte Carlo simulations of MDS. Calculating the unperturbed distance between elements with randomly-perturbed *positions* is the other common error-simulation approach.

There are unsatisfactory aspects to the Additive mechanism. For $d_{ij} << \sigma$, we find a significant possibility that $\delta_{ij} < 0$, not that we care, since δ_{ij} is never observed directly. Ramsay proposed a second, Multiplicative form of error:

$$\delta_{ij} = d_{ij}\, e^{e(\sigma / \surd 2)}$$

or $$\log(\delta_{ij}) = \log(d_{ij}) + e(\sigma / \surd 2).$$

This "log normal" error distribution appeals to common sense. δ_{ij} is always positive. The variance of δ_{ij} is proportional to d_{ij}, in line with the property of physical quantities, that the noise in measurements increases with the value being measured (1 mm can be measured more accurately than 1 km). A sort of scale independence of errors results: for two dyads, (i,j) and (k,l), with $d_{ij} > d_{kl}$, where the probability that $\delta_{ij} < \delta_{kl}$ is P, the same probability of error applies to the comparison between (s,t) and (u,v) where $d_{st} / d_{ij} = d_{uv} / d_{kl}$. Triadic comparisons for geometrically similar triangles in the configuration have the same error rates.

Incorporating a Multiplicative-error option to MTRIAD is straightforward. It is only necessary to use $\log(d_{ij})$ in place of d_{ij} in calculations of the Likelihood and the restorative forces. The effect is to devalue comparisons within large triangles, so that a judgment from a large-scale triad which conflict with X decreases the total Likelihood no more than a conflicting triad which is smaller but of the same proportions. This contrasts with ALSCAL, where d_{ij}^2 is used in Stress calculations, in effect ascribing more accuracy to comparisons, the greater the distances being compared.

6. FORMS OF RANKING DATA

To a child who's just been given a hammer, everything looks like a nail. The hammer is Johnson's formulation of MDS, as implemented in MTRIAD, with the useful features (individual variation models and MLE) introduced in Chapters 4 and 5. In this chapter I examine various non-triadic forms of proximity data to see how nail-like they are, i.e. whether they can be interpreted as pairwise comparisons between dissimilarities.

If data in two different formats are both reducible to dissimilarity comparisons, they can be combined (MTRIAD pays no attention to the method used to elicit any given comparison). This is my post-facto validation and justification for rashly performing such combinations.

Two intriguing forms are the methods of Sorting and Hierarchical Sorting, which warrant separate detailed scrutiny, so my attempt to bring them within the ambit of pairwise comparisons is postponed to Chapter 7.

Ranked Dyads

The most obvious case comprises experiments where comparisons between dissimilarities are made explicitly: a subject is presented with stimuli i, j, k, l and asked which pair is more similar: (i,j) or (k,l)? (or some equivalent question). With each judgment made independently, such tetradic data cry out for a Maximum Likelihood scaling as discussed in Chapter 5.

The number of pairwise comparisons increases as the fourth power of N. Bissett and Schneider [1992] discuss possibilities for omitting some or most of the comparisons, and perform Monte Carlo simulations to quantify how much the solution's accuracy is degraded by such omissions. For any i, j, some of the comparisons between (i,j) and other dyads can be missed out, while retaining others. Contrast this with directly rated dissimilarities, where the only option is to avoid assessing some δ_{ij} (chosen randomly, or with a balanced design, or interactively), which precludes all comparisons between (i,j) and other dyads. Triadic data form a subset of the set of tetradic comparisons, and clearly the triadic procedures are an example of selective omission.

This explicit method of pairwise comparisons is seldom used. In addition to Monte Carlo simulations, Takane [1978] applies his MLE analysis to tetradic comparisons between colour stimuli (plus Torgerson's triadic data for the same stimuli). In Schneider [1980], Schneider and Bissett [1981], Schneider, Parker and Stein [1974], the stimuli are sounds varying in volume in a single-dimensional configuration.

When the requirement of transitivity (consistency) is added to the completeness of pairwise comparisons, one has fully rank-ordered data. Chapter 2 described the special properties of full rank ordering: inequalities can be summarised by listing the dyads,

$$(i1,j2) \ll (i2,j2) \ll (i3,j3) \ll \ldots \ll (iN_d, jN_d)$$

Each pair is less similar than all the pairs left of them in the list, and more similar than all the pairs to their right.

In practice, this is exactly how the data are elicited. One prepares $N_d = N(N\text{-}1) / 2$ cards, one for each dyad: if, for instance, the stimuli are colours, each card consists of samples of the two colours, pasted to a neutral background. A subject sorts the cards into order of increasing dissimilarity. N is thus limited by the size of the surface on which the sorting is done, as well as by the subject's patience. Respondents, ranking pairs of countries [Klingberg, 1941] reported that more than the 21 dyads of $N = 7$ items would be unacceptable. Several decades later, more compliant subjects were willing to rank five times as many dyads, for $N = 15$ US states [Shepard & Chipman, 1970].

Shepard is the most notable practitioner of this method, applying it to states, digits [Shepard, Kilpatric & Cunningham, 1975], and colours [Shepard & Cooper, 1992]. In all cases, visual images were presented as one set of stimuli and the corresponding verbal label as another. Fillenbaum & Rapoport [1971] applied the procedure to colour names, kinship terms, and nouns for emotions (in Hebrew), in each case with $N = 15$. An early application of MDS in New Zealand [Clarke, 1976], with 9 names of fruit as stimuli, took full rank ordering for granted as being the *only* form of dissimilarity data.

All reported applications have used stimuli suitable for simultaneous presentation. All cards are out on the table at once, enabling the informants to perform the $O(N^4)$ comparisons between cards "in parallel" as they sort them into order.

Clearly such data can be analysed as a table of dissimilarities, by setting δ_{ij} equal to the ordinal position of (i,j) in the rank order: $\delta_{ij} = \Sigma\, \varepsilon_{ij.kl}$. The recovered $\Phi(x)$ (mapping d_{ij} onto δ_{ij}) must be monotonic, but that is all; the ipsatised nature of these values means that none of the quasi-metric arguments for restricting $\Phi(x)$ to exponential, or convex, or simple spline functions apply.

Is Maximum Likelihood Estimation appropriate? At best, it is an approximation. The comparisons are no longer independent. The nature of their dependence (which judgments are predetermined by previously-made ones) is unknown: they are not "directional data" [Takane & Carroll, 1981]. Multiple strategies are available to a subject ordering the cards, such as working inwards after initially picking the extremes of most and least similar, or grouping the cards into bands of similarity, prior to ranking them within each band. The experimental instructions sometimes recommend one strategy or another.

I turn now to variant forms of rank ordering which retain the transitivity feature while relaxing the requirement of completeness. The motivation is the usual one of lightening the burden on subjects.

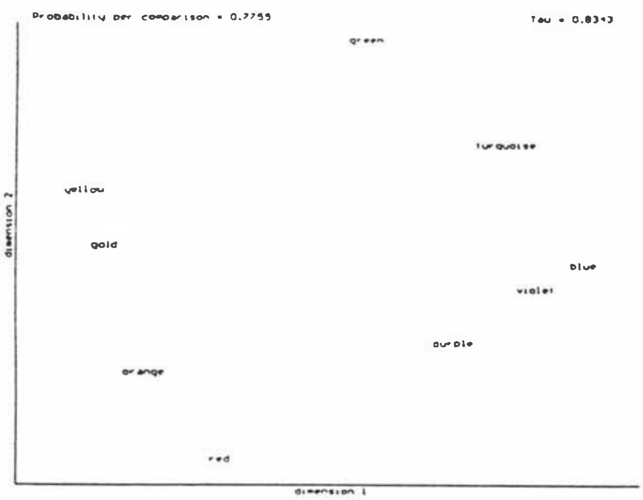
One can specifically reduce the number of comparisons between proximities of similar magnitude (arguably, the hardest ones to make) by asking the subject to sort the cards into groups of similar proximities. Barraclough [personal communication] performed a study where the groups have fixed sizes, in a manner similar to Stephenson's [1953] Q-sort procedure. There were 9 elements (colours), making a total of 36 cards. The two stimuli on each card were 2 cm squares, subtending about 3° when viewed at a comfortable distance. Subjects were requested to select the three cards (dyads) showing greatest similarity, and to place them in one pile; then to place the three least similar in a second pile; then to create two more piles containing the five next most similar and the five next most dissimilar; and so on. When finished, the cards are distributed in 7 piles, thus:

number in pile:	3	5	6	8	6	5	3	
(most similar)	1	2	3	4	5	6	7	(least)

The number of judgments is thereby reduced from 630 to 546: the 84 comparisons between similarities in the same pile remain unmade, imposing no constraints on the solution (when analysing the data, this is Kruskal's Primary treatment of ties). Analysing the comparisons with MTRIAD yields figure 6.1, a roughly circular arrangement of the stimuli. This replicates an earlier study of the same colours [Shepard & Cooper, 1992] which used complete ranking. In addition to colour *samples*, Shepard and Cooper scaled colour *concepts*. The coloured papers in that study could not be duplicated exactly so Barraclough used the nearest available Pantone sheets, having the following Munsell codes:

colour	Pantone	H V C (Munsell)	
(a) red	032U	4.6R	5.5 / 4.6
(b) orange	021U	9.9R	6.5 / 14.5
(c) gold	116U	8.4YR	8.0 / 10.9
(d) yellow	'yellow'	6.7Y	9.0 / 11.7
(e) green	354U	4.3G	5.8 / 9.3
(f) turquoise	313U	8.1B	5.2 / 7.8
(g) blue	072U	9.3PB	3.5 / 8.6
(h) violet	266U	2.2P	4.5 / 10.9
(i) purple	'purple'	0.4RP	5.1 / 12.8

Figure 6.1 Configuration for nine colours



Of the 17 participants in this experiment (14 male, 3 female), nine also provided hierarchical sorting data for the same stimuli. Four of the participants were colour-vision deficient in various degrees of severity.

Figure 6.1 was obtained using the INDSCAL model to rotate the axes to best fit. I will consider individual variations at more length in Chapter 7, combining the two kinds of data. For now, note that the experimental design is not ideal for detecting qualitative differences (colour-blindness), given the saturation and the large angular size of the stimuli.

The situation is subtly different when the number of cards in each pile is left up to subjects. It then becomes an application of the Method of Successive Intervals [Messick, 1956]. The distinction is particularly graphic in a series of studies where a binary sorting procedure was used, with 30 abstract geometrical shapes for stimuli [Silver, Landis & Messick, 1966]. Subjects were instructed to sort the 435 cards into two piles, "quite similar" pairs as opposed to "less similar" pairs. That done, they proceeded to subdivide each pile into two more, and so one, until 16 piles resulted. Another example was Landis, Silver, Jones and Messick [1967].

It would seem that at each stage of binary sorting, dyads are not being compared against each other, but rather, against some abstract threshold levels of dissimilarity. Label these thresholds Þ_1 to Þ_{15} ($\text{Þ}_1 < \text{Þ}_2 < \ldots < \text{Þ}_{14} < \text{Þ}_{15}$). At the first stage, the (i,j)-th dyad is placed in the "quite similar" pile if $\delta_{ij} < \text{Þ}_8$, and in the "less similar" pile if $\delta_{ij} > \text{Þ}_8$. In the first case, dyads are next compared against Þ_4; otherwise they are compared against Þ_{12}, and so on. A total of $4 \cdot 435$ comparisons are made. Without the benefit of a binary procedure, more judgments are required to sort dyads into groups, but unless the group sizes are fixed as in the Q-sort method, it can still be done by comparing dyads against threshold levels of similarity.

This is very close to the method of directly rating dyad dissimilarities. The POOC tried both methods [Coxon *et al*, 1975, Interviewing Instructions], asking some interviewees to assign numbers (1 to 9) to each pair of occupational titles, while others sorted cards, displaying the titles in pairs, into nine groups. The latter method takes advantage of the familiarity and ease of card-sorting tasks. I note that the Method of Successive Intervals restricts the possible stimuli – they must be suitable for displaying on cards – while probably raising consistency, since the cards are all present on the table as cross-checks and reminders of the threshold levels, cross-checks not available in a rating task, when a number is assigned to each dyad in isolation.

In Takane's ML analysis of successive interval data, the threshold levels between similarity bands emerge as part of the solution [Takane, 1981]. Foreshadowing MLE, part of the rationale for generalising successive interval methods from a single axis to the multi-dimensional case [Attneave, 1956; Messick, 1956] was a probabilistic argument. Like

Torgerson's treatment of triads [1952], the argument transforms non-metric responses to a form suitable for metric MDS.

The added complication of determining the thresholds is unncessary in Stress-minimising versions of MDS. This is easier to see in Kruskal's formulation. If two dyads are in adjacent intervals, $(i,j) \gg (k,l)$, but $d_{ij} < d_{kl}$ in the reconstructed configuration, then the lowest-stress value for the threshold between the two intervals is identical to the disparities, $Þ = ð_{ij} = ð_{kl} = (d_{ij} + d_{kl}) / 2$. Moreover, the forces acting on elements to minimise $\Sigma(ð_{ij} - d_{ij})^2$, bringing the distances into the same order as the dissimilarities, are identical to the forces which would act on elements to segregate distances within interval bounds correctly by minimising $\Sigma(Þ - d_{ij})^2$.

Conditional Ranking

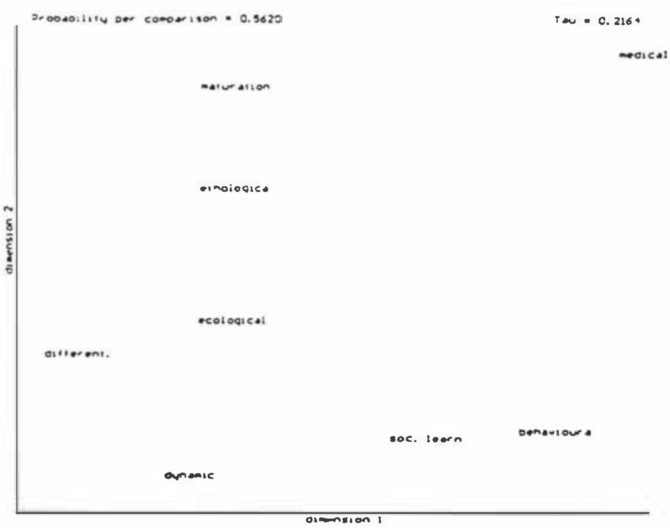
More radical reductions in the data requirements are brought about by using the method of *conditional ranking* (also known as rotating anchor points, group ranking, or multi-dimensional ranking [Klingberg, 1941]. This method involves choosing each element in turn as the anchor point, or hub, and ranking the remaining (N-1) elements in terms of increasing dissimilarity from that hub.

The ranking of the remaining elements could be done with pairwise (Thurstonian) comparisons, which would return us to the case of the Complete Method of Triads, with the triadic comparisons presented in a particular sequence. Normally a rank ordering is done [Klingberg, 1941; Jacobowitz, reported in Young, 1975; Kosslyn, Pick & Fariello, 1974]. Thus the same comparisons are implied as in the triadic case – the relationship between (i,j) and (k,l) is only known if the two dyads have a common element – but the requirement of transitivity within the conditional rankings takes away their independence (in Coombs' terms, each comparison provides fewer bits). For each anchor point there are (N-1)! possible rank orders for the remaining elements, a smaller number than the 2^{N-1} possible triadic responses.

As well as implying the same judgments, conditional rank-ordered and triadic data took equally long to provide, when Neidell [1972] compared the completion time between questionnaires worded in these two alternative form. However, the triadic form was seemingly perceived by its physician targets as harder, judging by the poorer response rate to mailed questionnaires. More systematic comparisons [Henry & Stumpf, 1975] found the triadic task taking between twice and three times as long to complete as conditional ranking, with N ranging from 7 to 15.

The Interactive Similarity Ordering procedure [Young, Null & Sarle, 1978] deserves a mention. ISO relies on a numerical sorting algorithm, 'mergesort', to order the (N-1) dissimilarities δ_{ij} (where i is constant), using multiple-choice questions of the form "which pair is closer?" The questions can be triadic in form. It is assumed that responses are error-free and transitive, allowing the relationships between uncompared dissimilarities to be inferred from them, minimising the number of explicit comparisons to be made. ISO is not limited to conditional data; it has an option for rank-ordering data fully.

Figure 6.2 Nine approaches to human development, conditionally ranked (M = 13)



Before moving on to the analytical implications for this conditional form of ranking, an illustration. The data are provided by Kirkland [personal communication]. At the end of a seminar on Human Development he asked the 14 students to conditionally rank the 9 schools of thought or approaches to human development that the course had covered. One data set was incomplete, but the other 13, combined, yield figure 6.2. Two dimensions in the configuration seemed to be enough. Examining the individual configurations before delivering them back to the participants gave the impression that much of the variation between them could be accounted for by dimensional weights, so I used the INDSCAL model. The orientation does not seem accidental: to reach it, the initial configuration rotated through about 30°. The dimensions are interpretable, with the theories arranged along the vertical axis from "social" (e.g. Social Learning, Behavioural) up to "physical" ones (Maturation), while the horizontal axis separates general theories of development (e.g. Medical) from more personal, "idiographic" ones (the Dynamic and Differentiation approaches).

Their conditional nature means there is no correspondence between an element's positions in the rankings from different anchor points. Depending on the configuration, element j may be most similar to i, while k is only second most similar to l, although $d_{ij} > d_{kl}$. Conditional rankings arrange dyads in a partially rank-ordered network. As with triadic data, distortions are introduced if one converts conditional rankings into a table of dissimilarities for scaling, approximating δ_{ij} as the sum of the ordinal position of (i,j) in the i-th and j-th conditional ranking, i.e.

$$\delta_{ij} \approx vc_{ij} = \sum_{k=1}^{N} (\varepsilon_{ij,ik} + \varepsilon_{ij,jk}) = \sum_{(k,l)} \varepsilon_{ij,kl}$$

It brings in constraints which have no basis in the data.

I have already mentioned the rigorous MLE treatment of conditional ranking data [Takane & Carroll, 1981]. Their analysis assumes that the ranking is, in their terms *directional*, i.e. it presupposes that subjects start by selecting the element most similar to the anchor point, progressing to the least similar, without revising the ranks at any stage.

Jacobowitz [Young, 1975] demonstrated that school-age subjects, the youngest being 6-year-olds, were capable of arranging semantic stimuli (words for colours, kinship, and body parts) in conditional rank orders.

Work by Indow and his collaborators with colour stimuli reveals the sizes of element sets this method can handle: $N = 21$ [Indow & Uchizono, 1960] and $N = 24$ [Indow & Kanazawa, 1960]. Their experimental design explicitly used the geometrical metaphor: samples of the (N-1) stimuli (colours) being rated for dissimilarity from an anchor point were presented concurrently, and slid back and forth by the subject until physical distances correspond to dissimilarities, yielding ratio-level data (other instances of ratio-level conditional rating appear in Arnold [1971], Hyman and Weil [1968], and Wish and Carroll [1974], the first detailing how response bias corrections should first be made to the ratings if they are to be treated as dissimilarities).

But N is never high enough, and the question arises of further increasing element sets by reducing the data requirements. Several variants of conditional ranking are in use.

One obvious option when there are enough subjects is sharing the anchor points amongst them. In ISO, each subject ranks elements' dissimilarities from c anchor points, c being set by the experimenter ($1 \leq c \leq N$). The artifical example in Chapter 4 of a reduced triadic data set was of this kind. Figures 4.3 and 4.4 remind us that although the distribution of the comparisons may preclude the recovery of a configuration from any single subject's responses, both the configuration and individual differences can be recovered if enough subjects contribute data.

This approach seems especially appropriate for investigating social networks, when the subjects are also the elements [Young, 1975]. Each subject is asked to rank all other subjects in terms of social distance. This covers the famous fraternity data [Nordlie, 1958], where 17 students, members of a newly-created fraternity, each listed the other 16 members in order of "favourable-ness of feeling" towards them. These rankings were elicited at weekly intervals over a four-month period, making them ideal material for analysis with the INDSCAL or Point-of-View models, to trace the evolution of the social structure. Tilstra [personal communication] applied this procedure to the nine departments within the Business Studies Faculty at Massey University: each departmental head rated the other departments, first in terms of actual closeness, then in terms of optimal (desirable) closeness.

In Shepard and Cermak [1973], subjects ranked 81 abstract forms for four levels of proximity ("very similar to the target", "next most similar to the target", "looks like or gives the same impression as the target", and the rest) from nine anchor points. The anchor points were the same for all 18 subjects. The configuration was known in this case, with the emphasis being to determine dimensional weights and the distance function (metric).
A second form of data incompleteness includes four variants, abbreviated to Order k/N, Order any/N, Pick k/N, Pick any/N [Coombs, 1964]. By the standards of MDS nomenclature, the

names are self-explanatory. The first two entail the subject considering each element in turn, ranking the k elements (out of the remaining N-1) closest to it. A value for k is fixed by the experimenter in the first variant, and left up to the subject's whim in the second.

The third and fourth variants are similar, except that the subject need only select the k elements closest to each anchor point, instead of ranking them. Thus, in their study of the perceived closeness of countries, Robinson and Hefner [1967] collected pick 3/17 data. White [1978] acquired pick 5/37 data, the 37 stimuli being personality-trait descriptors in the A'ara language. One interpretation and analysis open for such data would be as low-resolution conditional ratings: each element is either close to the anchor point, or less close.

The variants can be combined with the form of incompleteness described first. Thus, for one group of subjects in the Robinson-Hefner study, each subject only used nine of the 17 possible anchor points (different subsets of nine). Another example is Sampson's data [1968] for social distances within a monastery. 18 monastery members each ranked the other 17 on four relationships – Liking / Antagonism, Esteem / Disesteem, Influence / Negative influence,

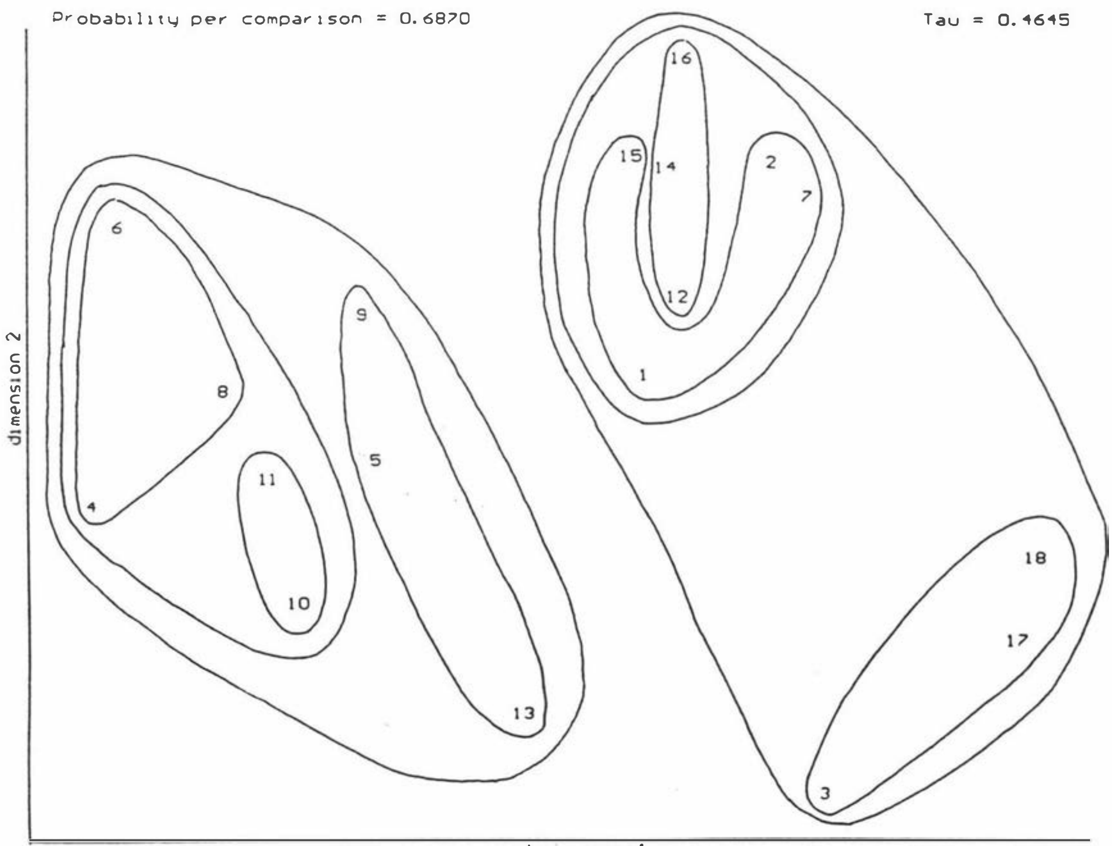


Figure 6.3 Monastery personal distances (time 4)

Praise / Blame: picking a first, second and third closest person for each relationship, and a first, second and third most distant. This leaves 11 persons in each ranking who are closer than the distant three but farther away than the closer three; the distances between them are unranked. The members repeated thisfor five time periods. Some made multiple first choices. Others declined to provide negative choices. Nevertheless, when these rankings are treated as low-resolution conditional distance ranks and scaled, the resulting two-dimensional configuration, figure 6.3, duplicates Sampson's social analysis (made on other grounds) of the members. The nested contours drawn around the points in figure 6.3 are the successively finer clusters obtained from the same data by applying the CONCOR algorithm [Breiger, Boorman & Arabie, 1975].

Thompson [1983] considered a variety of incomplete rankings. He collected conditional ranking data for a set of animal names, and observed the effects on the configuration of selectively including or omitting the ranks of stimuli in particular distance ranges from each anchor point.

The problem with picking or ranking only the nearest elements to each anchor point is the absence of rank-order information about larger dissimilarities. Graef and Spence found in Monte Carlo simulations [1979] that knowing the larger dissimilarities is vital for recovering the global structure of the configuration. If limited to short-range information, one can only reconstruct small portions of the "map", which are not guaranteed to fit together accurately. The problem becomes glaringly obvious when one deals with a highly clustered configuration, where Pick *k*/*N* or Order *k*/*N* responses constrain how the elements are arranged within each cluster, but give no indication about the relative positions of the clusters.

Rao and Katz [1971] evaluated these methods with simulated, noise-free data. They found that spurious dimensions intruded: setting $P = 2$, the dimensionality of the (known) configuration, left much variance unaccounted for in the vote-counted δ_{ij} and resulted in values of Stress in the range 0.22 to 0.31 (non-metric solutions). However, the two-dimensional solutions (ignoring individual differences and pooling data for 20 simulated subjects) gave reasonable recovery of the configura-tion. Under their analysis, the Order *k*/*N* and Order any/N methods displayed no superiority over the Pick methods. Individual dimensional weights were less accurately recovered than from fully ranked dissimilarity data. Robinson and Hefner [1967] summarise their subjects' responses in the form of a vote-counted proximity matrix, which, analysed with MDS, illustrates again the weakness of vote-counting. There are many dyads for which little is known (there are few constraints on their lengths), which thus receive small or zero proximities. MDS analysis takes this to indicate the rank order of their lengths and minimises Stress by positioning elements equally far apart, in a hyperspherical shell, occupying the full number of dimensions available. The dimensionality becomes high, perhaps spuriously so.

See also Green and Rao [1971], for a simulation of low-resolution ratings of dissimilarities, from which the familiar artifact of a circumplex configuration emerged.

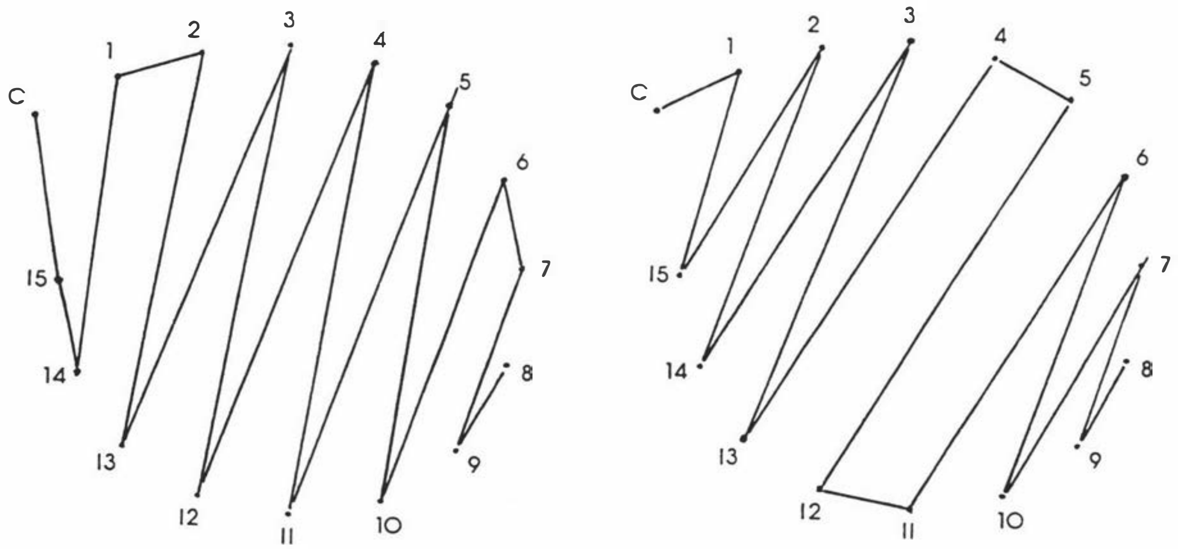
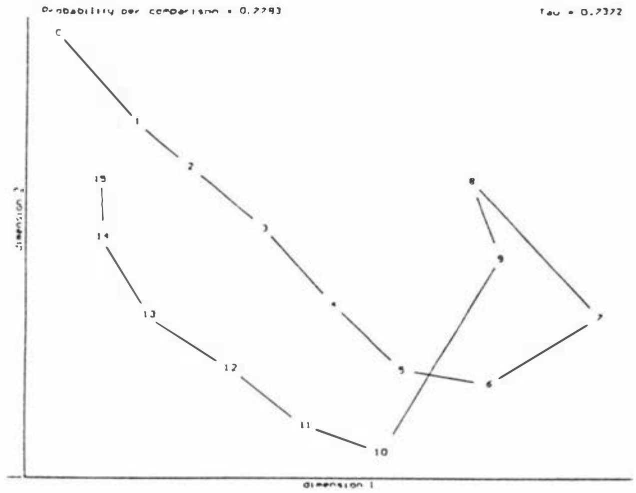


Figure 6.4: Aberrant closest-colour sequences for D15 stimuli (normal sequence is C, 1, 2, ..., 15)

An extreme form, the pick 1/N procedure, is used in the Farnsworth-Munsell D15 test of colour vision. To further summarise Chapter 4's outline of this test, one colour sample out of 16 (the stimuli) is the initial anchor, a, and from the remaining 15 stimuli, the subject chooses b, closest to a. This provides 14 dissimilarity comparisons: $\delta_{ab} < \delta_{ai}$, $(i \neq a, i \neq b)$. b now becomes the new anchor, and and c, the closest stimulus to it, is chosen from the remaining 14: providing another 13 comparisons. This process continues until the sequence of stimuli is finished and a total of 105 comparisons have been made – too few to reconstruct the configuration, even when data from several subjects with differing dimensional weights are combined (see figure 6.5). But Chapter 4 emphasised the point that when the circular configuration is given *a priori*, there is enough information to identify a subject's variations from normal. For someone with a significant colour-vision deficiency, the circular configuration becomes an ellipse in that person's private colour space, resulting in an aberrant closest-colour sequence (figure 6.4). However, the list of comparisons of the D15 test is too restricted for reliable assessment of subtle gradations of colour vision, for which the distorted circle is only slightly elliptical. The Farnsworth-Munsell 100-Hue test can be considered in the same light.

Figure 6.5 Solution for D15 responses ($M = 23$) treated as pick $1/N$ comparisons



A third form of data incompleteness has aspects in common with the first two. Partition the items into two groups, N_a hubs and N_s spokes. Only hubs serve as anchor points. For a given hub, only distances to spoke elements are ranked. Generally such experiments are cross-modal. Gregson [1966a, 1966b] champions cross-modal methods for situations where the sensory modality of the stimuli (taste, in Gregson's studies) makes direct comparisons between them difficult; they become hubs and are compared instead against easily-ranked visual analogues for the spokes.

In many cases the spoke items are categories or labels for the hubs. An example comes from the colour-naming experiment used in Boynton and Gordon's third experiment [1965], where the hubs were flashes of colour. Subjects opted for one or two colour names, chosen from a set of four choices (the spokes), which best matched each stimulus as it was briefly displayed. A more congenial way of collecting similar data is a wine-tasting [Winton, Ough & Singleton, 1974]. Participants select one or more wine varieties (spokes) as best descriptions of each wine (the hubs). Sometimes confusion matrices are treated in this way [van der Camp & Pols, 1971].

To describe this procedure as "conjoint scaling" invites confusion with other uses of the word "conjoint". It is also known as "unfolding" (generalising Coombs' term for preference scaling

[1964]). Coombs describes it as QIa data. I prefer the term "cross-modal" for the procedure, even though the hub and spoke elements are not necessarily of different modalities.
Thus, in a incomplete triadic design, Bechtel [1976] partitioned a set of Munsell colour chips into three "standards" and six "comparison" stimuli, the latter being compared, two at a time, for relative dissimilarity from each of the former (45 comparisons instead of 252). The distinction between standard and comparison stimuli is made frequently in applications of MDS to odours. Olfactory space (assuming the spatial model to be valid) is high-dimensional, and to map it in full requires many stimuli, too many for complete data to be conveniently collected, so Yoshida [1975] partitioned 72 odour stimuli into 32 test odours which were compared against 40 essential oils, requiring 1280 similarity ratings instead of 2556.

A final term for this kind of data is "off-diagonal", reflecting the partitioned nature of the dissimilarity matrix. In figure 6.6, sub-matrices 1 and 4 are unknown. Sub-matrix 3 = 2^T. Values are comparable only along rows of 2 (and columns of 3).

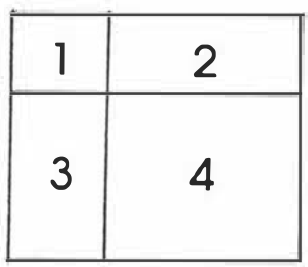


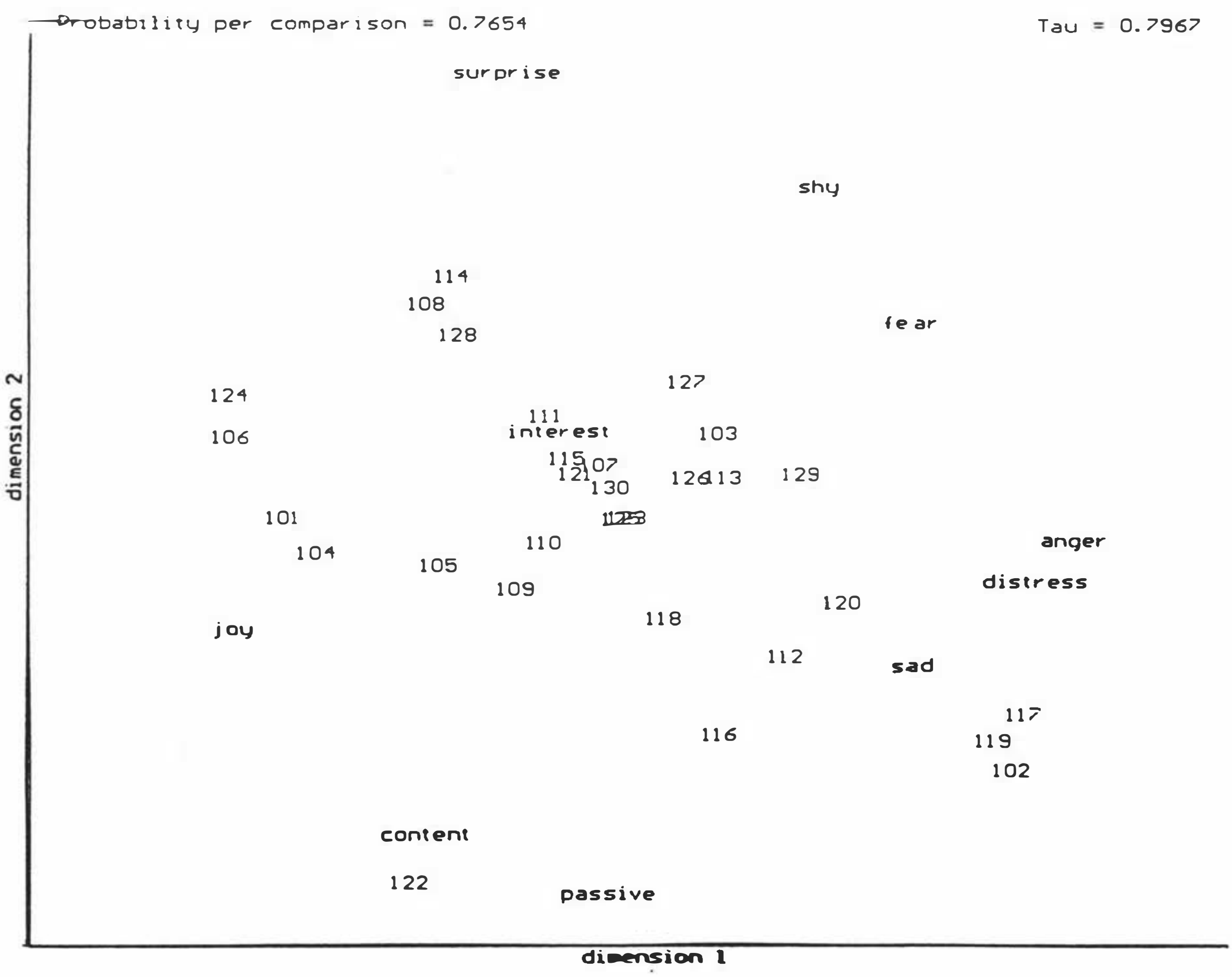
figure 6.6

A good example of cross-modal data appears in the documentation provided with the "I-FEEL" test. The I-FEEL pictures (*I*nfant *F*acial *E*xpressions of *E*motion from *L*ooking at *P*ictures) are supplied by the University of Colorado Health Sciences Center. They are 30 photographs of the faces of infants and young children, captured in various expressions, intended for use in a projective psychology test. Here I treat them as stimuli to be scaled. Chapter 7 will present two- and three-dimensional maps of these stimuli, created using hierarchical sorting data.

Part of the I-FEEL documentation is a table summarising people's reactions to the photographs. This table consists of the emotional terms ascribed to each photograph by 145 subjects, as percentages of total descriptions, in 13 columns: e.g. photograph '103' was described as displaying "Interest", 23.4% of the time; as "Sad", 14.5%; as "Distress", 13.8%; and so on. I ignored the column of "Other" responses, and also the columns of "Shame" and "Disgust" responses, which were hardly ever used. This left a 30-by-10 matrix of stimuli and the emotions attributed to them. For the analysis, I assumed that the labels and the photographed expressions could be represented as 40 points in an "emotion space", with the percentages indicating the proximity between a photograph (a hub) and the labels (spokes). The resulting two-dimensional map is figure 6.7, in which "Interest" lies at the centre of a rough circle formed by the other labels – an arrangement closely agreeing with maps such as Russell [1980].

The reader is invited to compare figure 6.7 with 7.12, a map derived from hierarchical data, and to admire the similarity between them. The labels allow the axes of the two diagrams to be identified as a vertical scale of Intensity or Activation (ranging from '122', a photograph of a sleeping child, to the extremes of fear and surprise) and a horizontal Pleasantness-Unpleasantness scale.

Figure 6.7 Two-dimensional configuration: distances between 30 I-FEEL faces and 10 emotion descriptions



Equivalent pick *k*/*N* procedures have been applied with personalities as hubs and personality traits as spokes [Rosenberg & Sedlak, 1972; Wing & Nelson, 1972]; pains and pain coping methods [Moore,1990]; varieties of pain and pain-descriptive phrases [Clark, Janal & Yang, 1984, Moore & Dworkin, 1988]; syndromes and symptoms. This is not to say that they were, or should be, analysed as conditional rank-orders.

The key requirement for cross-modal data to be analysed as conditional rank-orders is that hub and spoke elements must be represented within the same perceptual or semantic map. Nothing is known directly about the distances between pairs of hubs or pairs of spokes because such dyads never undergo comparison. This increases the opportunity for error in the reconstruction (the size of the confidence ellipsoids does not bear contemplating).

For a final, hypothetical example, consider using the Stroop effect to measure the proximity between colour names and actual colours. The Stroop effect is the slight increase in time taken to recognise a colour word when it is displayed in letters of a different colour. I speculate that displaying "RED" in *pink* letters would produce a smaller recognition delay than *green* lettering would, i.e. that the delay is somehow related to the distance between points representing colours and names in a conceptual / sensory colour space. Only the pressure of time, and its extreme irrelevance to the aims of this dissertation, prevent me from attempting such an experiment.

Cross-modal data have the same limitations as the use of scales. A dimension may be present amongst the stimuli, but it will not be detectable if it fails to separate the hubs. In Chapter 7, I present hierarchical sorting data which support the theory that emotions (as manifested in facial expressions) occupy a space of at least three dimensions. However, among the more common English nouns for emotions, a third dimension is weak or absent; two dimensions are enough to accommodate them. Thus, allowing a third dimension when analysing the table of emotions attributed to the I-Feel photographs does not produce any useful new distinctions. Instead, a two-level configuration is the result, with the labels all in one roughly planar group, displaced from another plane containing the stimuli.

Preferences

In his magisterial analysis of the different forms data can take, Coombs [1964] considered cross-modal conditional rank ordering as a more abstract case of preference data.

Coombs' *unfolding model* is a spatial representation for preferences (see also Roskam, [1968]). Points indicating the stimuli under investigation are positioned in a geometrical space, of low dimensionality, for I assume that relatively few forms of variation between stimuli suffice to explain the preferences of many subjects. I further assume that for each subject there is a combination of quantities which is optimal: no actual stimulus can be preferred more than it, and the more similar a stimulus is to the combination, the more it is preferred. Ideal points within the stimulus space represent these combinations.

Interpret the preference judgements as distance inequalities: if subject o prefers stimulus i out of the options i and j, x_i is closer than x_j to the ideal point x_o.

$$\delta_{oi} < \delta_{oj} \qquad \text{i.e. } \varepsilon_{oj,oi} = 1.$$

A list of preferences is clearly a conditional ranking, with the stimuli being ranked in order of their distances from a hub. So $N_a = M$, the number of subjects.

In many studies, each subject compares each pair of stimuli separately – "pairwise preferences" – creating triadic data [e.g. de Soete & Winsberg, 1993]. Ramsay [1980] points out that such data are richer in information and lead to more robust conclusions. In Delbeke's study [1968] on family composition preferences, the preferences of 84 university students were elicited as pairwise comparisons, but in the published form of the data these are summed and rank-ordered.

This seems like a good place to pause for an example. 12 blocks of Paulownia wood, identical in shape (20 x 85 x 285 mm) but varying in numerous aspects of the wood-grain, were ranked for preference by 43 subjects [Barraclough, personal communication]. A three-dimensional solution is adequate: figure W.1 (in Appendix W). The same subjects also provided similarity data for the blocks, following the hierarchical sorting procedure, so I will discuss the perceptual space in more detail and propose interpretations for the axes in the next chapter.

I preempt that discussion to the extent of noting here that the similarity-based configuration agrees with the preference solution. The postulated equivalence between preferences and distances is upheld. The results agree with Steinheiser's finding [1970], that preferences are more variable than similarity structures, reached after he elicited both forms of data from his subjects.

However, there remains the possibility that the two tasks involve different dimensional weights. There is no reason why the 'iso-preference' contours around a subject's ideal point should be circular or spherical. The subject can quite reasonably place a high emphasis on one or two particular features of the stimuli, while making a preference decision, and then proceed to equally weight the contributions the features make to the overall dissimilarity. It is even feasible that the features (i.e. dimensions) most salient to preferences do not show up in the similarity structure, where they are obscured by other features relevant to dissimilarities but not preferences. One can only hope that this is not the case for the stimuli in question. Fitting individual dimensional weights (the INDSCAL model) may help [Carroll, 1972].

There is a paradoxical flavour to this business of scaling preference data, though fortunately no real contradiction. I am assuming, firstly, that all the subjects derive their preferences from the same perceptual map (apart from random perturbations), and secondly, that their lists disagree, because of differing ideal points. Consistent preference lists, resulting from a single ideal point, would make life easier for market researchers and boring for everyone else, but scaling them would lead to degenerate solutions. If all ideal points have the same value c_r on the r-th dimension (i.e. they lie on a plane, $x_{or} = c_r$, $1 \leq o \leq M$), that dimension disappears from the solution – just as with other cross-modal situations.

In Chapter 7 I will combine the block preference data ($N = N_a + N_s$) with similarity data for the N_s wood blocks alone to derive a configuration more trustworthy than either data set could produce in isolation. It is worth stressing the instability of solutions for this kind of hub/spoke ranking data, with their absence of information about at least half of the distances, those corresponding to (hub, hub) and (spoke, spoke) dyads. Borg and Lingoes [1987] point out that there are always degenerate solutions with Stress = 0: e.g. all the ideal points at the centre of a hollow sphere of stimuli points. In the analysis of such data, one hopes that the algorithm will become caught in a meaningful local minimum rather than reach the degenerate global minimum.

I have noticed that MLE has many advantages over Stress minimising, much as in the case of incomplete triad designs. It seems better at avoiding both local minima and degenerate solutions. Figure W.1 is a MLE result.

One variety of data covered by this unfolding model is the Q-sort [Stephenson, 1953]. A universe of possible statements and opinions about some issue – Stephenson's 'concourse' – is represented as a multidimensional space. Embedded therein is a point for each statement, and an ideal point associated with each participant's own opinion. How far the o-th participant agrees with the i-th statement should correspond to the proximity of x_i to x_o.

I have applied MTRIAD to a number of Q-sort data sets: from Mrtek (34 statements about substance abuse, ranked by 85 medical students); Browne (40 suggestions for increasing faculty productivity, ranked by 11 academics); Kirkland (40 statements about future directions for Massey University, ranked by 13 Heads-of-Departments). The results were meaningful. In these examples, as in the majority of Q-sorts, N_s is too large for a complete ranking. Instead, the experimenter decides in advance on a restricted number of levels of agreement, and how many statements must go into each level (e.g. three statements "most agreed with", five "next most agreed with", and so on, down to five "next least agreed with" and three "least agreed with").

MDS on such data characterises the subject relationships: who agrees with whom. It also maps the concourse of statements itself. This is its advantage over the conventional factor-analysis treatment of Q-sorts. However, further work is in order. Independent, confirmatory maps of these concourses are needed, before proclaiming the superiority of MDS. One candidate for the cartography of concourses is the method of Sorting, the subject of Chapter 7.

It remains to specify the initial configuration, which is still an important factor in reaching the best final one. The incompleteness of preference data makes selecting $X^{(0)}$ difficult. The procedure used in MTRIAD is to position the N_s stimulus points first, applying PCO to a matrix of estimated inter-stimulus distances, which in turn are derived by correlating over hubs:

$$d_{ij}^{(0)} = \left(N_s \sum_o \mathrm{d}_{oi}\mathrm{d}_{oj} - \sum_o \mathrm{d}_{oi} \sum_o \mathrm{d}_{oj}\right) / \left(N_s \sum_o \mathrm{d}_{oi}^{\,2} - (\sum_o \mathrm{d}_{oi})^2\right)^{1/2} / \left(N_s \sum_o \mathrm{d}_{oj}^{\,2} - (\sum_o \mathrm{d}_{oj})^2\right)^{1/2}$$

Each of the N_a ideal points is then assigned the same position as that subject's most preferred stimulus (we cannot simply combine the $D^{(0)}$ matrix with a submatrix of hub-hub distances, estimated by correlating over spokes, plus the rank orders themselves for estimates of hub-spoke distances, since each sub-matrix has a different and unknown scale).

The Vector Model

The number of degrees of freedom in the ideal point model can be reduced by constraining the ideal points to lie outside the configuration of stimulus points – infinitely far out, in fact. True infinity is not necessary; it is sufficient for the ideal points to lie on a hypersphere at a fixed distance R, large enough that the vectors pointing from each of the stimuli to a given ideal point can be approximated by parallel lines (which would be springs, to return to that metaphor). Then the iso-preference contours for that ideal point are effectively planar.

This is the vector model for preferences, so-called because it is easier to treat the subjects' preferences as comparisons between vector products than between distances. It is therefore peripheral to the topic of this thesis, and will not be discussed at length; also, there is little to add to Roskam's discussion [1968]. The model might apply to qualities which have no finite optimum level; more is always better. Ramsay's example [1980] is sweetness, for which adults have ideal points – if a taste exceeds one's ideal sweetness, preference goes down – whereas children behave as if that optimum is an infinite concentration. In Coombs' taxonomy it is QIII data.

Represent the stimuli as a configuration of points, x_i, $1 \le i \le N_s$. I imagine vectors passing through the origin in the directions of the ideal points, at infinity, like knitting needles skewering a ball of wool; represent each by a unit-length vector y_o, $1 \le o \le N_a$. As R approaches infinity, d_{oi} approaches $R - x_i \cdot y_o$. Ignoring the constant R (since all the data are in the form of comparisons, i.e. differences between distances),

$$d_{oi} = - x_i \cdot y_o. \tag{6.1}$$

These "distances" are the projections of the i-th stimulus point onto the o-th "knitting needle". The best alignment of the vector (to be found in the course of MDS) is that which maximises the rank correlation between d_{oi} and the observed preferences. If one scans across the configuration in the direction of the o-th ideal point, one should encounter the stimuli in order of increasing preference, encountering the most preferred stimulus last: it should have the most negative projection onto y_o. Any violations of preference order contribute to Stress.

I adapt the familiar hill-descent algorithm to optimise the configuration, by inserting the new definition of "distance" into the Stress, and differentiating by x_i and y_o:

$$\begin{aligned} \mathrm{S_v}(X) &= \sum_{o,i,j} \varepsilon_{oij}\, \mathrm{H}(d_{oj} - d_{oi})\, (d_{oj} - d_{oi})^2 \Big/ \sum_{o,i,j} \varepsilon_{oij}\, (d_{oj} - d_{oi})^2 \\ &= \sum \varepsilon_{oij}\, \mathrm{H}((x_i - x_j) \cdot y_o)\, ((x_i - x_j) \cdot y_o)^2 \Big/ \sum \varepsilon_{oij}\, ((x_i - x_j) \cdot y_o)^2 \end{aligned} \tag{6.2}$$

$$\text{where } \varepsilon_{oij} = \varepsilon_{oi.oj} \quad \begin{cases} = 1 \text{ if } j \text{ preferred to } i \text{ by } o \\ = 0 \text{ otherwise.} \end{cases}$$

$$\partial \mathrm{S_v} / \partial x_i = 2 \sum_o y_o \sum_j \varepsilon_{oij}\, \Theta((x_i - x_j) \cdot y_o)^2 \,/\, \text{constant denominator} \tag{6.3a}$$

$$\partial \mathrm{S_v} / \partial y_o = 2 \sum_{i,j} (x_i - x_j)\, \varepsilon_{oij}\, \Theta((x_i - x_j) \cdot y_o)^2 \,/\, \text{constant denominator} \tag{6.3b}$$

Recalling that y_o must remain on the unit sphere, i.e. $\Sigma\, y_o^2 = R^2$, decompose $\partial \mathrm{S_v} / \partial y_o$ into components perpendicular and tangential to that sphere, and jettison the former (to put it another way, remove the component of $\partial \mathrm{S_v} / \partial y_o$ parallel to y_o). This task of keeping the magnitude of y_o constant could be avoided by expressing y_o in angular coordinates.

Some corollaries follow.

First, nothing is gained in this model by fitting individual dimensional weights. Planar iso-preference contours are still planar after transformation by a dimensional factor; it is not like transforming a circular contour into an ellipse. Any such transformations are equivalent to changes in y_o.

Similarly, nothing is gained by shifting to non-Euclidean geometries: the iso-preference contours remain planar. Perhaps one should think of the vector model as a special kind of geometry in its own right: the "infinity metric". In the limiting cases of the supremum and the city-block geometries, the available contours are restricted drastically: they must lie either perpendicular to an axis (in the first case) or diagonally (in the second), a reminder of Shepard's observation that these geometries have a propensity towards degenerate solutions.

Thirdly, zero-Stress solutions are always possible, just as in the ideal point model. In one instance of a degenerate solution in three dimensions, all the stimulus points are confined to a two-dimensional "pancake", with the preference vectors perpendicular to it.

MLE avoids this degeneracy. Again, the definition of L and the derivatives for maximising it follow from inserting the new "distance" function:

$$L_v = -\sum_{o,i,j} \varepsilon_{oij} \log(\exp(\tau\,(d_{oi} - d_{oj})) + 1)$$

$$= -\sum_{o,i,j} \varepsilon_{oij} \log(\exp(\tau\,(x_j - x_i) \cdot y_o) + 1) \qquad (6.4)$$

$$\partial L_v / \partial y_o = -\tau \sum_{i,j} \varepsilon_{oij}\,(x_j - x_i)\left(1 + \exp(-\tau\,(x_j - x_i) \cdot y_o)\right)^{-1} \qquad (6.5a)$$

$$\partial L_v / \partial x_i = -\tau \sum_{o} y_o \sum_{j} \varepsilon_{oij}\left(1 + \exp(-\tau\,(x_j - x_i) \cdot y_o)\right)^{-1} \qquad (6.5b)$$

$$\partial L_v / \partial \tau = -\sum_{o,i,j} \varepsilon_{oij}\,(x_j - x_i) \cdot y_o \left(1 + \exp(-\tau\,(x_j - x_i) \cdot y_o)\right)^{-1} \qquad (6.5c)$$

I note that nothing about the vector model restricts it to use with preferences. A feature which it shares with the ideal point model is that it can accommodate rankings of the stimuli on other scales, e.g. utility or beauty, known in the general case as dominance data: $\varepsilon_{mij} = 1$ if element i dominates j, i.e. is placed higher in the m-th rank order.

For a set of 13 facial-expression photographs, selected from the Lightfoot series, Cliff and Young [1968] obtained both pairwise dissimilarities and ratings of "Intensity" (presumably their quoted value of 7.65 for the Intensity of photograph 20, "physical exhaustion", is a misprint for *1.65*. I also assume that the point labeled 31 in their figure 1 is a misprint for *51*, "knows her plane will crash"). Using the ideal point model, they found that Intensity ratings were correlated with distance from a low-intensity "ideal point". It is probable, given that point's peripheral position in the configuration, that the vector model would have accounted for the ratings equally well.

Scale-based ratings are also covered, e.g. Semantic Differential scores, personal-construct ratings (the repertory grid procedure), or ratings on continuous or Likert-style 7-point scales. The crucial assumption is that each scale can be modelled by planar same-value contours in the stimulus space. Further candidates for this interpretation are the low-resolution pick any/N scales – syn-dromes, personality traits, etc. – which have only one same-value contour per scale. To bring them under the umbrella of Johnson's pairwise MDS, let $v(m,i)$ be the value on scale m of item i: then

$$\varepsilon_{mij}\begin{cases} = 1 \text{ if } v(m,i) > v(m,j). \\ = 0 \text{ otherwise.} \end{cases}$$

Recall the remarks by Bechtel and Coombs – previously mentioned in the context of triads – that preferences and scale values are both interpretable as forms of triadic data.

A common situation in MDS is where scale ratings and proximity data are both available. To this extent, the vector model is inside the scope of this thesis. I have already performed joint scalings on different proximity data sets, combining triads with dissimilarities, dissimilarities with ideal-point-modelled preferences. The concern now is with a *joint* analysis of scales and proximities [Ramsay, 1980, 1986], where the Stress-minimising (or Likelihood-maximising) forces which rearrange the configuration are responsive to both sets of data.

Here, $S_j(X) = S(X) + S_v(X)$
$L_j(X) = L(X) + L_v(X)$

This is distinct from both the External treatment of preferences, (also known as "conditional scaling of preferences", PROFIT for Property Fitting, or Prefmap for Preference Mapping [Carroll, 1972]), where the vectors (or ideal points) are adjusted to optimise their fit within a pre-established configuration of stimuli which is not altered in the process (often with the aim of helping interpret it), and the Internal treatment, or preference scaling, as just seen.

A case in point is the woodblock stimulus set of appendix W. As well as providing the preference and similarity data mentioned before, the blocks were rated by wood-technologist judges. The vectors y_m for the 12 scales with the highest correlations with the configuration of figure W.1 were added to the model, with a corresponding L_v term, which introduced additional terms in ∇L. Initial values $y_m^{(0)}$ come from least-squares regression lines fitting the scale values $v(m,i)$ to $X^{(0)}$.

After scaling, the result was figure W.4. The configuration is much the same, but the presence of the scales allows its axes to be identified.

Configurations like W.4, where one dimension accounts for less variation than the others, creates an irritating artifact when combined with poorly-fitting scales. In a recurrence of the degenerate "pancake" solutions, Stress is minimised when such scales are aligned parallel to the minor axis. Likelihood is maximised as well, i.e. the effect persists under MLE. It can be difficult to distinguish scales which label the minor axis from those which are aligned with it as an artifact, though the latter have higher Stress contributions.

It was possible to test the validity of the woodblock configuration and the scales, because an additional 12 stimuli were on hand. Though they had not been used in similarity experiments, they had been rated on the same scales by the same judges. We could incorporate them in a 24-block configuration, . in which the extra 12 points (woodblocks) were defined by scale values alone, predicting where they would lie if proximity data were available as well. For joint analysis, it is not necessary for the element sets in the different data sets to be identical, only that they overlap. Later, preferences and similarities did indeed become available for the additional blocks, producing figure W.5, which verifies the prediction and vindicates the chain of assumptions.

24 stimuli being too many to be conveniently ranked by preference, they were partitioned into two sets, of 12 stimuli each, to be ranked separately by the subject and scaled separately, with

two ideal points. This partitioning was arbitrary and different for each subject. The closeness of the two ideal points per subject is a source of confidence in the model.

Curiosa

This illustrates the multiple aspects of supplementary data. Data sets which were originally collected for scaling purposes can also function as tests of one's basic assumptions. Thus, my confidence that triads are a form of dissimilarity comparison was bolstered when the triadic data for occupational titles [Coxon *et al*, 1975] arranged them in the same configuration as the dissimilarity ratings.

In an experiment with a set of animal names [Henley, 1969], triadic and other forms of data were collected, providing similar convergent validity when scaled. With the assumptions validated, the data sets can be combined in a joint analysis. Examples of this will appear in the next chapter.

But conversely, data collected with the intention of testing a configuration's validity can also be used to refine it. I mention these collection procedures for the sake of completeness, without advocating their use for scaling purposes, though some may have a role to play, e.g. to augment sparse data sets (pick k/N, etc). They serve as a reminder that no single data-collection procedure is ideal for all purposes, and that some situations may require a combination of complementary procedures.

An example is Isaac's "odd-one-out" test. The subject is given one stimulus, i, and asked to select another three j, k, l such that the initial stimulus will stand out as the least similar of the four. It is predicted, if the spatial model is applicable and the configuration is correct, that the test is one of distance inequalities: the distances d_{ij}, d_{ik}, d_{il} should all be greater than the three d_{jk}, d_{kl}, d_{jl}.

In one example I have already discussed, baby-cry triads [Lyons, Kirkland, Castle & Lowoko, 1991], the additional data are the estimates from subjects H and I of their confidence in their judgments. There is a correlation of 0.674 between the estimates and the differences between the reconstructions of the distances being compared, $|d_{ij} - d_{ik}|$, in figure 3.2. With faith in the reconstructed map thereby bolstered (not to mention faith in the estimates themselves), one can include them in the construction of the map, though the changes from doing so are minor.

Rumelhart and Abrahamson [1973] used a method of analogies to test configurations in semantic space [McAdams & Cunible, 1992; Wessel, 1979 applied this method to timbral space]. The questions are multiple-choice: "*a* is to *b*, as *c* is to (*t*, *u*, ...)?"[1] In an extension of the Henley [1969] work, Rips, Shoben & Smith [1973] had the good fortune to find two semantic domains (bird names; mammal names), both two-dimensional, with homologous dimensions: "size", and "domesticated-tame-feral-predatory". This enabled them to perform "cross-modal" analogy tests, taking *a* and *b* from one domain and *c*, *t*, *u* from the other. They found the distances in the recovered configurations to be better predictors of analogy choices than than the raw, "unpurified" dissimilarities were.

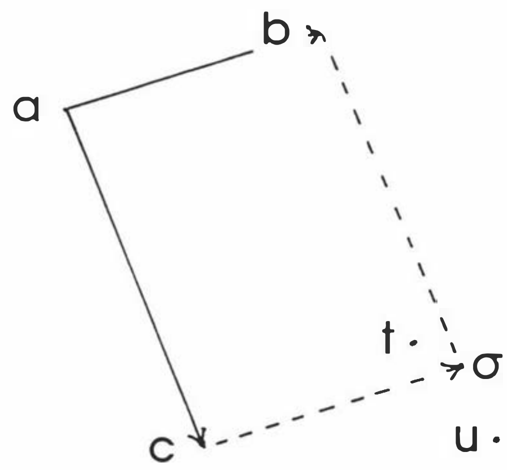


Figure 6.8 Parallelogram of vectors locating σ

If the spatial model is correct, the analogy creates a *virtual stimulus*, σ (see figure 6.8):

$$x_\sigma = x_c + (x_b - x_a)$$

[1] Kant considered analogies with three empirical, observable terms while the candidates for the fourth term are metaphysical or mystical: "*nicht etwa, eine unvollkommene Ähnlichkeit zweier Dinge, sondern eine vollkommene Ähnlichkeit zweier Verhältnisse zwischen ganz unähnlichen Dingen*" [Prolegomena to Every Future System of Metaphysics that May Ever Arise in the Way of a A Science]

$(x_b - x_a)$ being the vector that translates stimulus point x_a onto x_b. The distances between x_σ and the possible answers x_t, x_u, etc. are being compared: by choosing t, the subject asserts $\delta_{\sigma t} \ll \delta_{\sigma u}$, etc. Any disagreements between the actual and the predicted choice contribute to Stress, which we can minimise, as before, by calculating gradients, i.e. corrective forces acting on x_t, x_u, and x_σ.

$\partial S / \partial x_t$, $\partial S / \partial x_u$ are straightforward. As for $\partial S / \partial x_\sigma$, note that changing the positions for any of the three items a, b, c affects the virtual stimulus. The force on σ resolves into corrective forces on a, b, c, since $\partial x_b / \partial x_\sigma = \partial x_c / \partial x_\sigma = -\partial x_a / \partial x_\sigma = 1$.

$$\begin{aligned}\partial(\text{raw Stress}) / \partial x_{bp} &= \partial(\text{raw Stress}) / \partial x_{cp} = -\partial(\text{raw Stress}) / \partial x_{ap} = \partial(\text{raw Stress}) / \partial x_{\sigma p} \\ &= \sum_t \left(\partial(\text{raw Stress}) / \partial d_{\sigma t}\right) \left(\partial d_{\sigma t} / \partial x_{\sigma p}\right)\end{aligned}$$

This kind of data is conceivably useful for scaling a set of elements where N is less than the threshold for reliable recovery of the metric and configuration; one could augment the available stimuli with virtual stimuli σ, adding landmarks to the map. A simpler way to create virtual stimuli using fewer points (tetradic data, rather than pentadic) involves multiple-choice questions of the form "Which stimulus, c or d, is closest to being midway between a and b?" Then $x_\sigma = (x_a + x_b) / 2$ (putting it another way, $x_b - x_\sigma = x_\sigma - x_a$).

Carroll and Chang [1972] considered both forms of data, analogies and mid-points, as vector equations, and wrote a program, SIMULES, for exploiting them as material for MDS (relying on ratio-level data). Another paper which looks at performing MDS using vectors [Gordon, Jupp & Byrne, 1989] deals with a slightly different form of data, where the *direction* from one point to another is known, but not the distance between them.

The analogy-task nature of a third example of vector equation (relevant to the research with facial expressions to be described in Chapter 7) was not recognised by the researchers [Russell & Fehr, 1987]. Here, photographs of facial expressions were presented in pairs. The first expression displaced the second within "emotion space", affecting the way it was perceived. For instance, if the face in the first photograph was smiling, a neutral, relatively expressionless face presented second looked sad in comparison. Subjects indicated the location of the second stimulus x_2' in a two-dimensional emotion space by rating it on scales or by picking the closest verbal label (cross-modal data). Thus x_2' is a virtual stimulus:

$$x_2' = x_2 + 0.41\,(x_0 - x_1)$$

where x_0 is the origin of the coordinate system (corresponding to neutrality or lack of emotion) and x_1 and x_2 locate the first and second stimuli as observed in isolation. The factor 0.41 was found by Russell and Fehr empirically. They found the displacement to be additive: with two stimuli, x_1 and x_3, to displace x_2,

$$x_2' = x_2 + 0.41\,(x_0 - x_1) + 0.41\,(x_0 - x_3)$$

Russell has argued [Russell 1980] in favour of interpreting emotion space with a system of polar coordinates (in which an angle and a distance from the origin suffice to specify an expression or emotion), as an alternative to an x-y system of axes. Mathematically the coordinate systems are interchangeable. However, these results indicate that the x-y system is a better model of subjects' internal representations of expressions. To account for their observations, Russell and Fehr are reduced to converting the polar coordinates for a stimulus into x-y coordinates and performing the vector sum before converting back to the polar representation. Their subjects did not combine the angular and radial components of two expressions directly.

Another case for choosing between rival geometrical representations is Colour. Holding luminance constant, colour space can be interpreted with polar coordinates (the radial coordinate being saturation, while the angular coordinate is hue: the standard colour wheel), or with a pair of axes at right angles (i.e. the CIE system). Polar coordinates seem intuitively more natural (to someone with colour-normal vision), but x-y coordinates are a closer representation of the early levels of colour processing where opponent processes are at work, as shown by individual variations, subjects' versions of colour space displaying compression along one or other axis. It is perhaps worth conducting analogy tests to see whether subjects conform to a single coordinate system in their internal representation of colours, and if so, what that system is.

7. SORTING AND HIERARCHICAL SORTING

One aspect of the sorting and the hierarchical sorting methods is important enough to warrant mentioning before going on to describe the nature and the analysis of the data. This is the fact that disagreements between the subjects' responses are not only expected, but necessary. Given complete unanimity among informants, the only solutions possible are degenerate. This reliance on disagreement links the sorting methods to Torgerson's analysis of triadic data. It also serves as a pleasant metaphor for the conflicting views required in science if any progress is to be made.

For the INDSCAL model to be applied, or any other model of individual variation, we must somehow sift out systematic differences in a subject's responses from the random inter-replication errors required in order to scale them. It will be interesting to see if this is possible.

Sorting

The method of sorting involves a simple partition of *N* stimuli into groups or piles ("partitioning" having the sense of defining an equivalence relation: the groups do not overlap but between them include all the stimuli). Usually the items are printed on cards: in the form of samples if they are perceptual, or words, pictures or labels for conceptual stimuli. The subjects are asked to sort these cards into piles, on the basis of similarity, so that stimuli which are most similar are in groups together. The exact wording varies. The criteria for assessing similarity, the number of stimuli in each group, and usually the number of groups are all left up to the subject. Hence it is also known as unconstrained sorting, or free sorting (F-sorts). Fillenbaum and Rapoport [1971] use the term "direct grouping".

I stress the simplicity and ease of administration of the method, which have led to its wide-spread use, despite being amongst the most recent of the procedures I have mentioned; it was first used by linguists, to explore semantic "spaces" in the late 1960s [Anglin, 1970; Clark, 1968; Miller, 1969; Steinberg, 1967]. It has been used with children as young as 3-year-olds [Russell & Bullock, 1985, 1986], and with other cultures [Berlin, Breedlove & Raven, 1968;

Lutz, 1982; Russell, 1983; Russell, Lewicka & Niit, 1989]. As the flip side of this simplicity, the quantity of data obtained per subject is not enough for individual solutions. Miller [1969] recommends recruiting at least 20 judges, but more are usual. With only 13 judges [Lutz, 1982], the reliability of the results is questionable.

Contributing to the popularity of the method is its capacity to scale large *N*, essential in any attempt to map a culture's cognitive structure (as indicated by persons' use of words). 100 stimuli are not unknown [Kraus, Schild & Hodge, 1978; Miller, 1971]. With the objective of comparing cognitive structures across cultures, Church and Katigbak [1989] asked 15 Filipino students to sort 74 (English) personality descriptors, while another 15 sorted 89 descriptors (in Tagalog), and 15 sorted the combined, bilingual set of 163. The record (in the English-language literature) is perhaps 176 pain-descriptive words [Verkes, van der Kloot; & van der Meij, 1989].

The sorting task as just described, where the elements are physically rearranged on a flat surface, is unsuitable for administration *en masse*. For eliciting data from a room of people simultaneously, the triadic method in combination with screen-projected stimuli is more appropriate [e.g. Gladstones, 1962a].

The nature of the task restricts it to stimuli that can be presented in parallel. There are exceptions to this: sorting has been successfully applied to tactile textures, presented sequentially [Hollins, Faldowski, Rao & Young, 1993], and to odours [Lawless, 1989; MacRae, Howgate & Geelhoed, 1990; MacRae, Rawcliffe, Howgate & Geelhoed, 1992; Paddick, 1978].

A search of the literature uncovered the following sorting applications. They are listed in no particular order.

- "verbs of having" (English) [Fillenbaum & Rapoport, 1971; Takane, 1980]
- "verbs of breaking" [Hojo, 1993]
- prepositions [Clark, 1968; Fillenbaum & Rapoport, 1971]
- countries [Wish & Carroll, 1974]
- body parts [Miller, 1969]

- kinship terms [Rosenberg & Kim, 1975]
- role terms [Burton & Romney, 1975]
- statements about occupations [Coxon & Jones, 1978, 1979a, 1979b]
- occupations [Burton, 1972, 1975; Coxon & Jones; Kraus, Schild & Hodge, 1978]
- forms of interpersonal behaviour [Burton, 1975]
- strategies of getting one's own way [van der Kloot & van Herke, 1991]
- values [Jones, Sensenig & Ashmore, 1978]
- stereotypical adjectives, and often-stereotyped ethnicities [Jones & Ashmore, 1973]
- adjectives of "good" and "bad" [Fillenbaum & Rapoport, 1971]
- media explanations for riots [Schmidt, 1972]
- influential figures in psychology [Rosenberg & Gara, 1983]
- pain behaviours [Turk, Wack & Kerns, 1985; Vlaeyen, van Eek, Groenman & Schuerman, 1987]
- pain coping strategies [Wack & Turk, 1984]
- pain descriptors [Clark, Janal & Carroll, 1989; Morley, 1989; Reading, Everitt & Sledmere, 1982; Torgerson & Melzack, 1970; Verkes, van der Kloot & van der Meij, 1989]
- pain varieties [Moore & Dworkin, 1988; Moore, Miller, Weinstein, Dworkin & Liou, 1986]
- personality traits [Church & Katigbat, 1989; Miller, 1974; Rosenberg, Nelson & Vivekananthan, 1968; van der Kloot & van Herk, 1991]
- environments [Ward, 1977]
- hand-signs for letters in American Sign Language (with two criteria for sorting; similarity of visual appearance and of three-dimensional conformation) [Richards & Hanson, 1985]
- facial expressions [Nummenmaa, 1988, 1990; Russell & Bullock, 1985, 1986; Russell, Lewicka & Niit, 1989; Stringer, 1967]
- words for emotions [Lutz, 1982; Russell, 1980, 1983; Russell, Lewicka & Niit, 1989]
- visual textures [Harvey & Gervais, 1978]
- plant varieties (species of gourd) [Berlin, Breedlove & Raven, 1968]
- names of colours [Fillenbaum & Rapoport, 1971; Takane, 1980].

The subjects can be entire cultures, for when one word in a particular language extends to label more than one item (colours, emotions, animal species, prepositions), those items have been sorted. MDS can be as easy as looking up words in multi-lingual lexicons.

I note that the sorting method is also applicable to social networks. In that case it is a question of analysing observations rather than experimental outcomes. See, for instance, the Struhsaker observations of vervet monkey sleeping groups [Arabie & Boorman, 1973]: instead of stimuli, the items were individual monkeys, partitioned into groups every night as they settled down to sleep in separate trees. Multiple nights, rather than subjects, provided the replications. Similar data can be obtained wherever a group of primates is accessible, e.g. who shares a table at an office cafeteria, or which Supreme Court judges vote together in judgments.

The list is not intended to be exhaustive. Its size is a result of a determined search through the literature for papers which quote the original, unprocessed sorts themselves. The size also suggests that there is a potential market for alternative scaling procedures for sorting data.

Note also the large number of studies applying the sorting method to pain-related fields. Their results have implications for both the diagnosis and the treatment of pain. This potential for affecting people's lives, unusual in psychology, gives a certain degree of immediacy to any improvements in analysis. One powerful diagnostic tool, the McGill Pain Questionnaire, was developed by Melzack and Torgerson [1970] from sorting data. The MPQ sets out lists of pain-descriptive words. The subject locates the pain afflicting him or her in the same mulitdimensional space as the words by picking which ones describe it (pick any/*N* data).

The absence of the sense of hearing from the list is disappointing, given my specific interest in baby crying. A trial sorting of the 15 complex tones of the fourth experiment of Chapter 3, using a version of the procedure of Faldowski *et al*, was a failure. It became apparent that the groups would be formed on the basis of verbalised representations of the stimuli: the subjects' memories of the tones themselves were too ephemeral to classify like with like directly. Bricker and Pruzansky [1970] and Pruzansky [1969] are often cited as sorting experiments, but the abstracts of the papers make it clear that subjects provided two-dimensional representations of the dissimilarities, by arranging pegs in a pegboard (linked to a computer). This intriguing procedure ("spatial sorting" is Miller's [1976] term for it) is worth reviving with a mouse replacing the pegboard interface [Goldstone, 1994], but it contains more information than sorting in the sense used here.

My proposal for a dissimilarity-comparison approach for scaling sorting data will be easier to explain in the context of hierarchical sorting data (H-sorts). I turn to this procedure now.

As before, the procedure begins with N stimuli, usually in the form of cards which can be spread out up a table or floor. The subject's task is to arrange them into progressively fewer groups, selecting the two most similar groups at each stage and merging them into a single group, a process repeated until one group remains. This process is hierarchical in that a group inherits all the members of the two groups which were merged to make it. There is no way for two items which are first grouped together at stage g to be separated again in some later stage. The result can be shown as a hierarchy, or tree diagram, or dendrogram, as in figure 7.1(a).

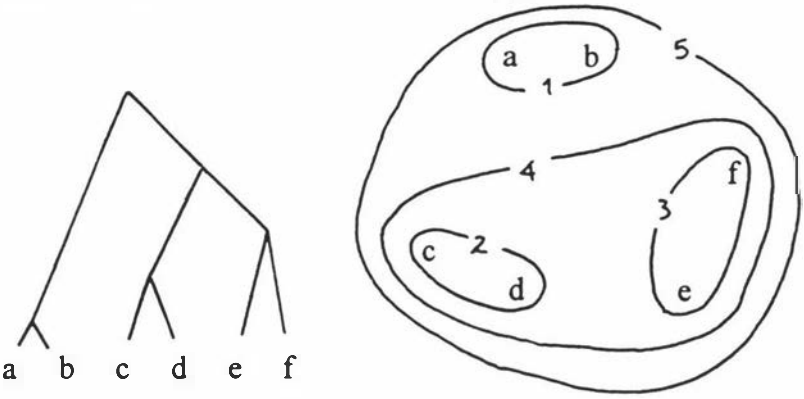
It is helpful to consider the initial situation as consisting of N piles, each containing a single stimulus. In what follows, I use upper-case letters $I, J, K, \ldots$ to label the piles or groups, to distinguish them from $i, j, k, \ldots$, the lower-case items, elements or stimuli. However, to suggest such an interpretation to subjects is a recipe for confusion and looks of bafflement. For the first group-merging stage, them, the subject is instructed to select the two most similar stimuli, i and j, and to place them in a pile I'.

For the second stage, there are two possible cases the subject must consider: the most similar pair of piles may be (a) two single cards j and k; or (b) I' and a single stimulus. At subsequent stages, as well as joining single cards and adding cards to piles, a third option appears: the most similar pair may be (c) two piles, each comprised of more than one card. The challenge for the experimenter is to explain these multiple interpretations of "most similar pair", without giving in to the subjects' understandable requests for an exact definition of "most similar".

With each stage decrementing the number of piles by one, N-1 stages are required, the last being trivial.

Figure 7.1 Different representations of an H-sort

(a) Dendrogram (b) nested clusters (c) matrix, C_m



	a	b	c	d	e	f
a	-	1	5	5	5	5
b	-	-	5	5	5	5
c	-	-	-	2	4	4
d	-	-	-	-	4	4
e	-	-	-	-	-	3
f	-	-	-	-	-	-

Figure 7.1 depicts several ways of presenting the same information: as a dendrogram, in panel (a); as nested 'contour lines' in panel (b). The same information is present when one records explicitly the subject's sequence of merges. Each time two piles are merged, it is not necessary to repeat all their members – a single member is enough. So figure 7.1 is equivalent to

a,b / c,d / e,f / c,e

Coxon *et al* used a less linear way of recording the merging sequence:

$$({}_5({}_1\, a\ b)^1\ ({}_4({}_2\, c\ d)^2\ ({}_3\, e\ f)^3)^4)^5$$

Here the indexed parentheses $({}_1$ and $)^1$ enclose the first two stimuli grouped together, $({}_2$ and $)^2$ enclose the next merging stage, and so on. Effectively this is figure 7.1(b), in one dimension.

A final form is cumbersome but suitable for computer input (it is the format MTRIAD expects):

(((((a b)))) (((c d)) (e f)))

The stage at which two groups are merged is indicated by the number of nested parentheses enclosing them. This format can be obtained from figure 7.1(a) by traversing the tree, writing '(' every time one descends a level and ')' on every ascent.

In the POOC research [Coxon *et al*, 1975] H-sorting was conducted with 16 occupational labels. Coxon and Jones testify as to the method's convenience and acceptance among their informants. Fillenbaum and Rapoport used it with a variety of semantic items: 24 names of colours, 15 kinship terms, 30 verbs of "judging", 20 adjectives each of "good" and "bad", 16 pronouns, 15 Hebrew words for emotions, 29 prepositions, 18 conjunctions, 29 verbs of "having". It seems odd that the method has not been used more widely, since it returns substantially more information than F-sorting, for little extra labour on the participants' part. However, Barraclough has applied H-sorting to a variety of stimulus sets, and I am fortunate in having had access to her unpublished data, as well as my own, for the examples which follow.

Filllenbaum and Rapoport's subjects took about one hour to H-sort sets of 29 and 30 items. For less motivated subjects or more complex stimuli, a value of 16 for N is approaching the practical limit. An option for scaling larger N is to ask each subject to H-sort a subset of items, the subsets overlapping to make up the complete set: in other words, an incomplete design. Fillenbaum and Rapoport used such a design to scale 40 adjectives of "good" and "bad". They asked their subjects to H-sort the 20 "good" and 20 "bad" adjectives separately, then to pair up the subsets to indicate how the separate configurations overlapped.

The smallest number of informants in the experiments to be described was 12. More are required with incomplete designs.

Coxon and Jones used a process comparable to vote-counting to convert the data into a table of distance estimates, for MDS:

$$\delta_{ij} \approx 1 - e_{ij}, \qquad \text{where } e_{ij} = 1/M \sum_{m} c_{m,ij} \tag{7.1}$$

Here $c_{m,ij}$ is the stage when i and j were first grouped together by the m-th subject (see figure 7.1(c)). With some of their data sets (conjunctions; verbs of judging), Fillenbaum and Rapoport asked subjects to rate the dissimilarity level at each merging of piles; one can use these values for $c_{m,ij}$ instead of the merging stage. However, I argue now that a better analysis is possible.

What are subjects going through in their minds when they choose the closest two groups, at the g-th stage, out of the $(N\text{-}g)\ (N\text{-}g\text{+}1) / 2$ possible pairs? How are they defining the perceived dissimilarity between groups I and J, δ_{IJ} ? Of sundry definitions which come to mind, two stand out:

$$\text{(a) } \delta_{IJ} = \max_{\substack{i \in I \\ j \in J}} \delta_{ij} \qquad\qquad \text{(b) } \delta_{IJ} = \min_{\substack{i \in I \\ j \in J}} \delta_{ij} \qquad\qquad (7.2)$$

These are simply the two Hierarchical Clustering Methods (HCAs) described by Johnson [1967], the Maximal (or Diameter) and the Minimum (or Link) algorithms respectively. The appealing feature of the two forms is that they reduce inter-group distances to dyadic dissimilarities, allowing observations such as "the m-th informant merged piles I and J at stage g, judging them to be closer than I and K, J and K, *et cetera*" to be interpreted as $(i,j) \ll (i,k)$, $(i,j) \ll (j,k)$, ...

Thus H-sorts become another special case of distance comparison data, with this peculiar feature, that the data do not explicitly identify the dyads being compared. It is necessary to deduce them.[1]

The items i,j for which $\delta_{ij} = \delta_{IJ}$ are defined by (7.2). This is of little help since it defines unknown quantities in terms of other unknowns, the dissimilarities δ_{IJ} not being part of the data either. I fall back on the central assumption of MDS: that $\delta_{ij} = \Phi(d_{ij})$ + an error term (where $\Phi(x)$ is monotonic). Given a provisional reconstruction $X^{(t)}$ of the configuration, and assuming that the subject uses either (a) or (b) of (7.2) – MTRIAD offers both options, but for the sake of concreteness let us settle on the Link algorithm, (7.2(b)) – one can deduce which dyads determined the inter-group dissimilarities at a given stage g in a particular subject's sequence of merging, by applying (7.2) to the reconstructed distances d_{ij}. These are precisely the dyads which must be moved together, or further apart, to bring $X^{(t+1)}$ into closer accordance with the observations.

[1]Note that in Fillenbaum and Rapoport's version of the H-sort procedure, these dyads *are* identified. Subjects were prompted at each stage for the dyad which provided the actual bridge between the piles they merged, and directed to follow the Link algorithm for clustering, resulting in a minimum spanning tree from each subject. See also Rapoport [1967]. Though providing more information, this variant has the disadvantage that it forces subjects to follow a pile-merging strategy which may not be the one they would use naturally.

I write $ij(I,J)$ to indicate the pair of elements $i \in I$ and $j \in J$ having minimal (maximal) reconstructed separation, d_{ij}. Thus $d_{IJ} = d_{ij}$. There is no need to search through the members of the group after every merging stage to find the defining dyads, because of another pleasant feature of the Link and Diameter algorithms, their recursive definition. The notation is more cumbersome than the arithmetic:

when the subject merges piles I and J into I', (7.3)
for all other piles K (where $K \neq I$, $K \neq J$),
if $d_{IK} < d_{JK}$
then $d_{I'K} = d_{IK}$, $ij(I',K) = ij(I,K)$
else $d_{I'K} = d_{JK}$, $ij(I',K) = ij(J,K)$.

This is a good point to mention a third feature of the two HCAs. They are extreme forms. If they lead to similar results, I will be spared the task of writing software to handle more complex algorithms (centroid, U-statistic, etc) wherein the distance between two groups I and J is not localisable to a single dyad but rather emerges as a communal property of *all* the dyads (i,j) (where $i \in I$ and $j \in J$). We will find that the choice of HCA for reconstruction makes little difference, both in Monte Carlo simulations where the particular HCA used to generate the data is known, and in the case of real-world data.

It is worth emphasising that the HCA is not being applied to the configuration to cluster the elements: the subject has already done that. It is being used within the constraints of the subject's clustering, to deduce the inter-group distances and the dyads determining those distances.

These preliminary remarks have provided practically all the components necessary for defining a Stress for H-sort data.

$$S_m = \sum_{g=1}^{N-1} \sum_{\substack{I,J \in \wp(m,g) \\ K,L \in \wp(m,g)}} \varepsilon_{m.g.IJ.KL} \, H(d_{ij(I,J)} - d_{kl(K,L)}) \, (d_{ij(I,J)} - d_{kl(K,L)})^2 \qquad (7.4)$$

Here $\wp(m,g)$ is subject m's partition into $(N+1-g)$ groups prior to the g-th stage of merging, $\wp(m,g) = \{ I_1, I_2, I_3, \ldots I_{N+1-g} \}$,

and $\varepsilon_{m.g.IJ.KL} \begin{cases} = 1, \text{ if the } m\text{-th subject selects } I \text{ and } J \text{ as closest at stage } g; \\ = 0 \text{ if } I \text{ and } J \text{ are not selected, or if } (I,J) = (K,L). \end{cases}$

This has the desired property that $S_m = 0$ if the configuration correctly predicts which piles are closest at each stage in the sequence. The worst possible reconstruction, creating the largest possible Stress contribution, is one where the pair of piles which are closest according to the data (i.e. which the subject merges) have maximal d_{IJ}. To keep the total Stress ≤ 1, write

$$\text{raw Stress} = \sum_{m=1}^{M} S_m$$

$$S = \text{raw Stress} / \text{normalisation factor}$$

$$= \sum_m S_m \Big/ \sum_m \sum_g \sum_{IJ,KL \,\in\, \wp(m,g)} \varepsilon_{m,g,IJ,KL} \left(d_{ij(I,J)} - d_{kl(K,L)}\right)^2$$

Note how Stress is dominated by comparisons made at early stages, presumably between short distances: at stage g there are $(N+1-g)(N-g)/2 - 1$ comparisons between the pair the subject chooses as closest, and all other pairs, each contributing a term to the denominator and potentially to the numerator. At the later stages where larger dissimilarities are compared, there are fewer piles and less effect on Stress. Hence, one cannot rely on accurately recovering the global structure of a mental map from H-sort data alone.

$$\partial(\text{raw Stress}) / \partial x_{ip} = \sum_j \left(\partial d_{ij} / \partial x_{ip}\right) \left(\partial(\text{raw Stress}) / \partial d_{ij}\right)$$

$$= \sum_j \left(\partial d_{ij} / \partial x_{ip}\right) \sum_g \sum_{IJ \,\in\, \wp(m,g)} \left(\partial d_{ij(IJ)} / \partial d_{ij}\right) \left(\partial(\text{raw Stress}) / \partial d_{ij(IJ)}\right)$$

$$= \sum_j (x_{ip} - x_{jp})/d_{ij} \sum_g \sum_{IJ \,\in\, \wp(m,g)} \delta_{m,g,ijIJ} \sum_{KL \,\in\, \wp(m,g)} \varepsilon_{m,g,IJ,KL}\, \Theta\left(d_{ij(I,J)} - d_{kl(K,L)}\right) \qquad (7.5)$$

$$\text{where } \delta_{m,g,ij,IJ} \begin{cases} = 1 \text{ if } (i,j) = ij(I,J) \text{ for } i,j \text{ and } I,J \in \wp(m,g) \\ = 0 \text{ otherwise} \end{cases}$$

MDS by minimising Stress turns out to be inadequate. Succumbing to the form of degeneracy described in Chapter 5, Stress is minimised by configuration of items spaced at equal intervals. The scarcity of large-dissimilarity comparisons allows this artifact to emerge.

As before, we have recourse to the Maximum Likelihood method. Let $D_{IJ,KL} = d_{ij(I,J)} - d_{kl(K,L)}$.

$$L = \sum L_m \qquad \text{where } L_m = -\sum_{g=1}^{N-1} \sum_{IJ,KL \,\in\, \wp(m,g)} \varepsilon_{m,g,IJ,KL} \log(\exp(D_{IJ,KL}\,\tau) + 1) \qquad (7.6)$$

$$\partial L / \partial x_{ip} = -\tau \sum_m \sum_j (x_{ip} - x_{jp})/d_{ij} \sum_g \sum_{IJ \in \wp(m,g)} \delta_{m.g.ijIJ} \sum_{KL \in \wp(m,g)} \varepsilon_{m.g.IJ.KL} \{1 + \exp(D_{IJ.KL}\tau)\}^{-1}$$

It remains to specify the starting configuration in this procedure. $X^{(0)}$ comes from Principal Coordinates Analysis of a table of estimated dissimilarities, E, obtained as in (7.1).

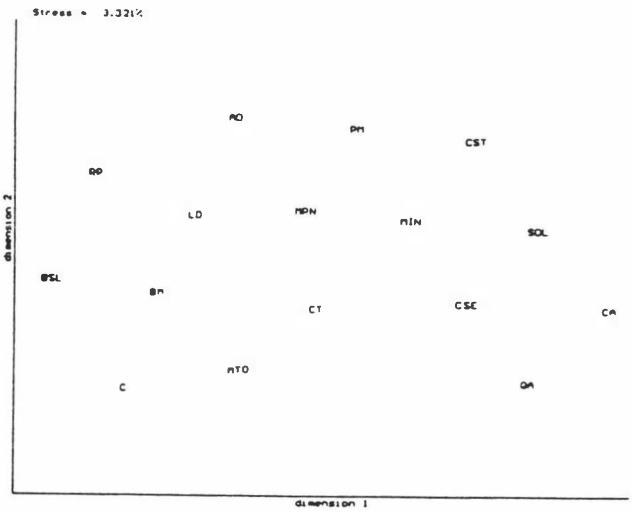


Figure 7.3 Stress-minimising configuration

Figure 7.3 illustrates the equal-spacing artifact, by showing a Stress-minimising analysis of sorting data with $M = 103$ for $N = 16$ occupational titles (using the Link form of hierarchical clustering). Figures 7.4(a) and 7.4(b) maximise Likelihood for the Link and Diameter forms of hierarchical clustering respectively. Both are very close to figure 5.3, the results of scaling dissimilarity data from 286 informants for the same items. The close resemblance between (a) and (b) of figure 7.4 supports my earlier assertion that the assumption of one hierarchical clustering strategy as opposed to another has little effect on the outcome.

Many of the subjects in the H-sort tests I have conducted personally (for colours, kinship terms, and facial expressions) volunteered the information that they were adding a stimulus card to a pile, or combining two piles, by taking the two *closest* stimuli from two piles to determine the distance between them. From now on, only results from the Link strategy will be reported.

This reanalysis of the occupations data does not demonstrate any clear-cut advantage over the conventional "vote-counting" approach of scaling the co-occurence matrix (7.1) (see figure U.3.2 in Coxon and Jones [1979b]). I collected data to test for any difference.

Stimuli were 16 sample swatches of Dulux® housepaint, selected to cover the widest possible range in hue and saturation, while minimising the variation in brightness. The selection was

Figure 7.4 (a) MLE reconstructed-dyad analysis for occupational-title H-sorts (M = 103), using Link HCA

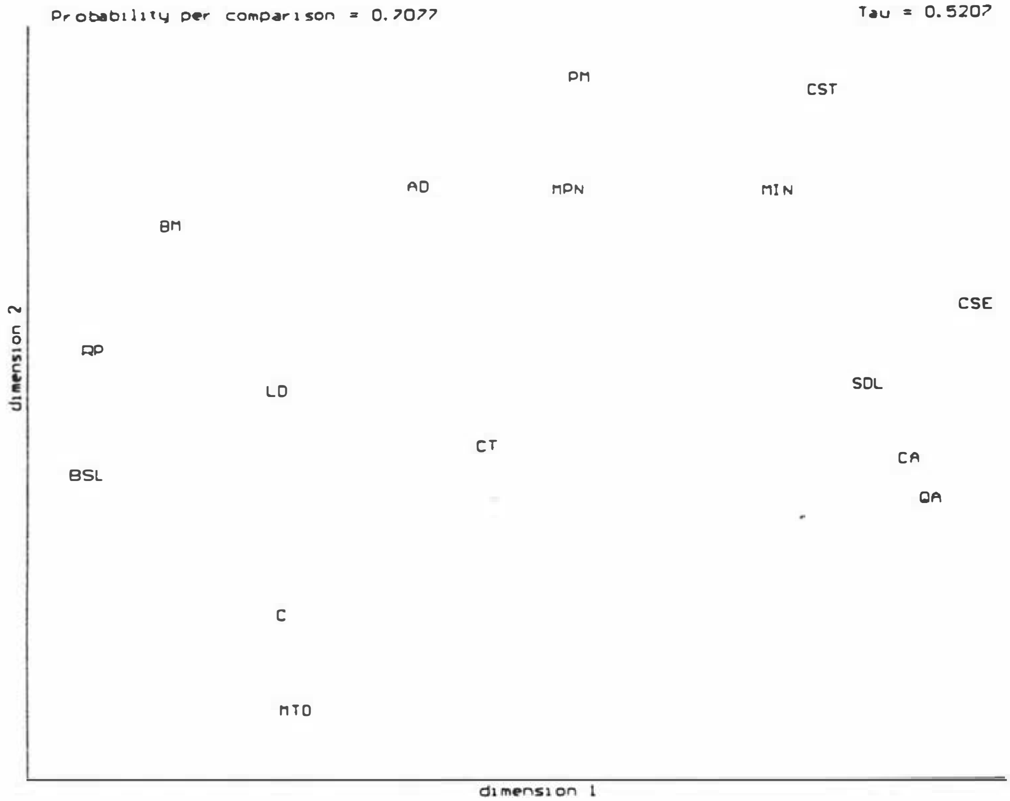
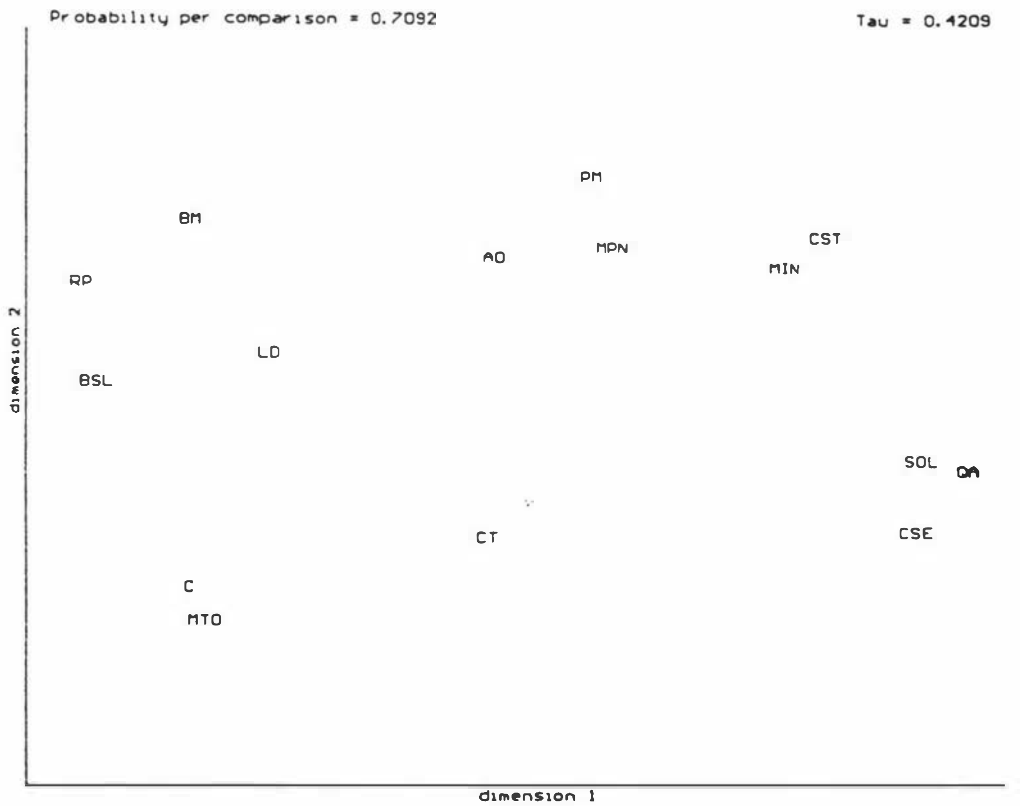


Figure 7.4 (b) MLE reconstructed-dyad analysis for occupational-title H-sorts, using Diameter HCA.



necessarily a compromise between these incompatible goals; brightness could be kept only approximately constant, while parts of the colour circle were not represented (there is a lack of vivid orange and violet housepaints). All swatches were of size 35 mm by 70 mm. Numbers for these colours are arbitrary, but are listed with their Munsell notations in Table 7.1. Figure 7.5(a) gives approximate locations for the stimuli in the (Hue,Chroma) colour plane.

The H-sorts from 9 subjects, in the Coxon-Jones format, are listed in Table 7.2. Barraclough collected a further 10 H-sorts, making a total of 19 subjects, (6 F, 13 M). Their data were analysed by this Method of Reconstructed Dyads, producing 7.5(c). Figure 7.5(b) is the result of applying the "vote-counting" procedure to the same data: a plethora of high δ_{ij} values results in a circumplex solution, with the unsaturated tans, pinks and beiges (included among the stimuli to ensure that the true configuration is *not* a circumplex) flung to peripheral positions as if by centrifugal force.

Table 7.1. Numbers and Munsell codes (H V / C) of 16 colours in the Dulux paint range.

No.	Munsell	No.	Munsell	No.	Munsell	No.	Munsell
4	7.5Y 9.1 / 6.2	47	2.2Y 8.4 / 6.6	68	9.2YR 7.0 / 6.2	112	9.0R 7.4 / 7.2
156	3.0R 7.3 / 8.2	173	2.0R 6.0 / 5.3	195	5.5RP 6.1 / 3.3	261	5.9PB 6.4 / 3.5
288	1.8PB 7.1 / 5.8	372	5.8G 6.1 / 2.2	376	5.8G 7.0 / 5.5	399	0.1G 7.8 / 6.0
416	4.4GY7.4 / 5.1	421	9.2YR6.7 / 2.6	443	4.1R 7.0 / 2.3	459	9.3B 6.8 / 1.3

Table 7.2 H-sorts from nine subjects for the colours (numbering them 1 to 16).

$$(_{13}(_{6}9(_{1}8~16)^{1})^{6}(_{12}4~(_{11}5~(_{5}(_{3}6~7)^{3}~15)^{5})^{11})^{12})^{13}(_{14}(_{10}14~(_{7}1~(_{2}3~2)^{2})^{7})^{10}(_{9}10~(_{8}11~(_{4}12~13)^{4})^{8})^{9})^{14}$$

$$(_{13}(_{4}(_{3}8~9)^{3}~16)^{4}~(_{7}10~(_{6}12~(_{5}11~13)^{5})^{6})^{7})^{13}~(_{14}(_{11}14~(_{9}(_{8}2~3)^{8}~1)^{9})^{11}(_{12}(_{10}5~4)^{10}~(_{2}6~(_{1}15~7)^{1})^{2})^{12})^{14}$$

$$(_{13}(_{2}4~5)^{2}~(_{12}(_{6}6~7)^{6}~(_{5}15~14)^{5})^{12})^{13}~(_{14}(_{10}9~(_{7}10~(_{3}8~16)^{3})^{7})^{10}~(_{11}(_{9}13~(_{8}11~12)^{8})^{9}~(_{4}1(_{1}2~3)^{1})^{4})^{11})^{14}$$

$$(_{13}(_{8}5~4)^{8}~(_{12}(_{2}16~(_{1}8~9)^{1})^{2}~(_{4}(_{3}6~15)^{3}~7)^{4})^{12})^{13}~(_{14}(_{11}(_{9}10~11)_{9}~(_{10}13~12)^{10})^{11}~(_{7}14~(_{6}(_{5}3~2)^{5}~1)^{6})^{7})^{14}$$

$$(_{8}(_{2}1~2)^{2}~3)^{8}~(_{14}(_{13}(_{7}4~5)^{7}~(_{11}14~(_{5}6~(_{1}15~7)^{1})^{5})^{11})^{13}~(_{12}(_{9}9~(_{3}8~16)^{3})^{9}~(_{10}10(_{6}13~(_{4}11~12)^{4})^{6})^{10})^{12})^{14}$$

$$(_{14}(_{12}(_{5}2~1)^{5}~(_{7}3~14)^{7})^{12}~(_{11}(_{6}5~4)^{6}~(_{4}7~6)^{4})^{11})^{14}~(_{13}(_{10}(_{3}12~13)^{3}~(_{8}10~11)^{8})^{10}~(_{9}(_{1}8~9)^{1}~(_{2}16~15)^{2})^{9})^{13}$$

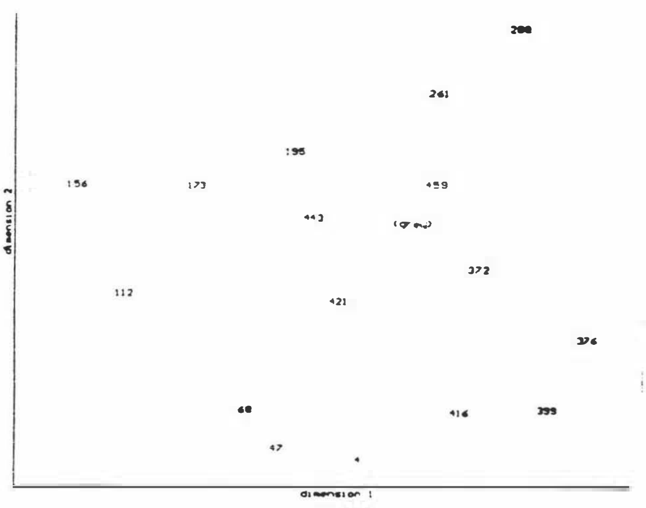
$$(_{7}3~(_{5}1~2)^{5})^{7}~(_{14}(_{12}(_{10}(_{8}(_{6}9~(_{1}8~16)^{1})^{6}~10)^{8}~11)^{10}~(_{3}13~12)^{3})^{12}~(_{13}(_{9}4~5)^{9}~(_{11}(_{4}6~(_{2}7~15)^{2})^{4}~14)^{11})^{13})^{14}$$

$$(_{14}(_{5}5~4)^{5}~(_{12}(_{6}6~(_{1}7~15)^{1})^{6}~(_{8}16~(_{7}8~9)^{7})^{8})^{12})^{14}~(_{13}(_{11}14~(_{3}(_{2}1~2)^{2}~3)^{3})^{11}(_{10}10~(_{9}11~(_{4}12~13)^{4})^{9})^{10})^{13}$$

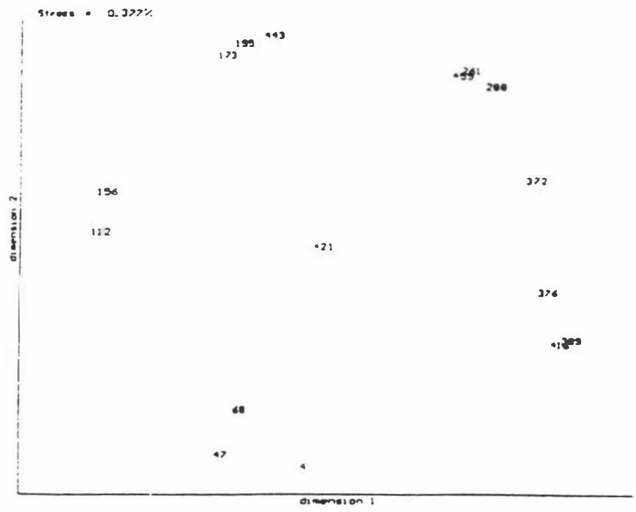
$$(_{7}3~(_{2}1~2)^{2})^{7}~(_{14}4~(_{13}5~(_{12}(_{6}15~(_{4}6~7)^{4})^{6}~(_{11}(_{5}9~(_{3}16~8)^{3})^{5}~(_{10}(_{8}10~14)^{8}~(_{9}11~(_{1}12~13)^{1})^{9})^{10})^{11})^{12})^{13})^{14}$$

In my next example, the true configuration *is* a circumplex. The D15 test for colour perception uses two stimulus sets of 16 colour samples each. Within each set the saturation is constant (as is luminance), leaving only hue to vary. Thus the 16 stimuli in each set are

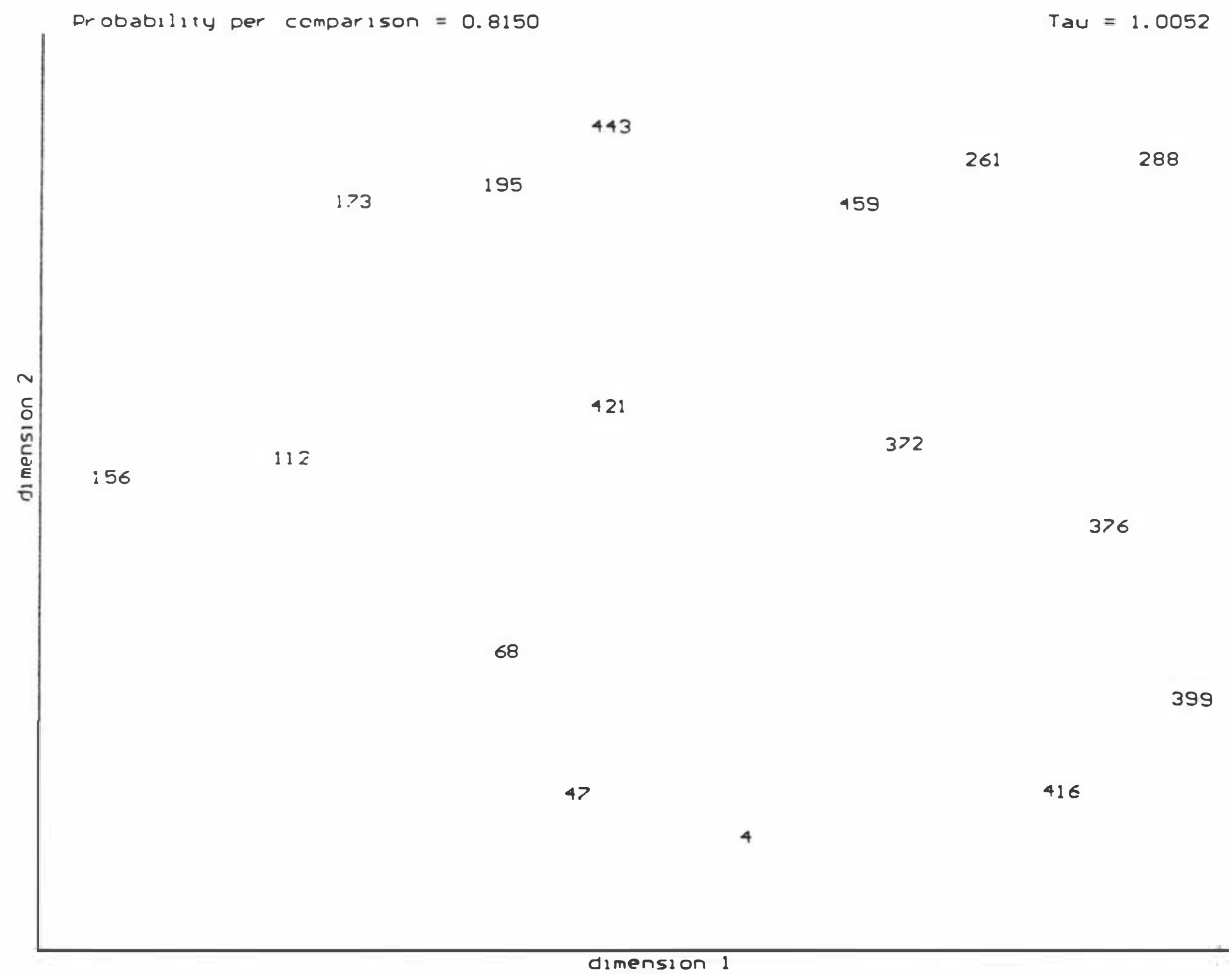
Figure 7.5(a) 16 stimuli in the Munsell plane



(b) Vote-counting analysis for 19 H-sorts



(c) Reconstructed-dyad analysis for 19 H-Sorts of 16 colour stimuli (using MLE and the Link HCA)



arranged in a circle in colour space. Barraclough obtained a total of 37 hierarchical sorts of the two stimulus sets, many subjects sorting both sets.

To an observer with normal colour vision, the stimuli are equally spaced around the appropriate (saturated or unsaturated) circle, far enough apart (compared to $1/\tau$) that the two stimuli closest to a given stimulus are its nearest neighbours in the circle: there is little chance that the nearest-but-one will be picked as "closest". This means, in practice, that in the course of hierarchically sorting the stimuli, a pile grows by merging with piles or single stimuli adjacent to it in the circle.

For the purposes of scaling H-sort data, the configuration is effectively one-dimensional. There is only local information, relating to each element's position relative to its two immediate neighbours, with none of the global information which would be required to recover the circularity. The data could equally likely have come from a string of stimuli bent into an ellipse or oblong. The outcome of applying MTRIAD to Barraclough's D15 data, in two dimensions, is figure 7.6, resembling a dropped necklace.

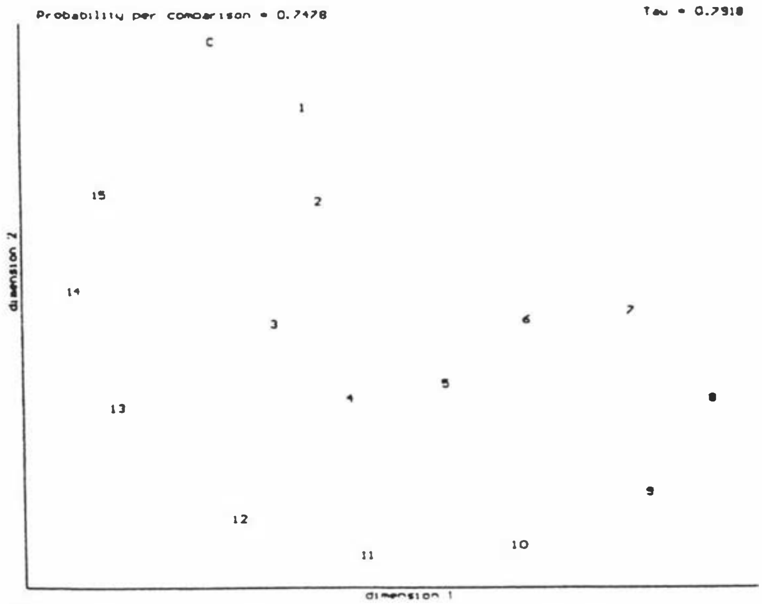


Figure 7.6 Solution for D-15 H-sort data ($M = 37$)

This demonstrates that some configurations lend themselves to reconstruction more readily than others. As the number of close neighbour elements j, k, l, ... to i increases (such that $|d_{ij} - d_{ik}| < 1/\tau$, $|d_{ij} - d_{il}| < 1/\tau$, ...), so does the number of distance comparisons involving i. Hierarchical sorting is sensitive to the positioning of landmarks within the perceptual space to be surveyed.

As a preliminary test of whether individual variations are discernible on H-sort data, I *imposed* the known circular configuration on the H-sorts, by scaling them jointly with a circulant distance matrix (recall from Chapter 4 that when the configuration is known in advance, less data are required to determine individual variations). In this situation the interesting aspect of the solution is the extent that individuals disagree with it. Individual Likelihoods are listed as table 7.3. Subject initials are in upper-case, for H-sorts using the Saturated stimulus set, and lower-case for the Unsaturated set. Note the low Likelihoods for subjects JH, MR, JW, MM (i.e. jh, mr, jw), etc., who have been diagnosed independently as colour-vision deficient (protanopic and deutanopic), and who sorted the stimuli in abnormal sequences.

Table 7.3 Likelihoods for 37 H-sorts.

S	Likelihd.	S	Likelihd.	S	Likelihd	S	Likelihd.	S	Likelihd.
CG	0.7976	DK	0.8284	PC	0.8275	rb	0.8157	eb	0.8222
SU	0.8279	JW	0.6511	RT	0.8125	ky	0.8255	mb	0.7283
PM	0.8155	RE	0.8148	MR	0.5302	bt	0.8307	pc	0.7869
PW	0.8140	JH	0.5603	jk	0.8010	mc	0.7134	rt	0.6182
GJ	0.8327	MM	0.5418	be	0.7917	wm	0.8135	mr	0.4983
KY	0.8048	EB	0.8365	ec	0.5816	jw	0.6008		
MC	0.7271	CH	0.5269	ae	0.8208	re	0.7819		
WM	0.8327	MB	0.8309	em	0.6483	jh	0.2915		

Is there room for the INDSCAL or the Points-of-View models in the analysis of H-sorting? Before pursuing the question, I summarise some previous work on distinguishing individual differences between hierarchical groupings.

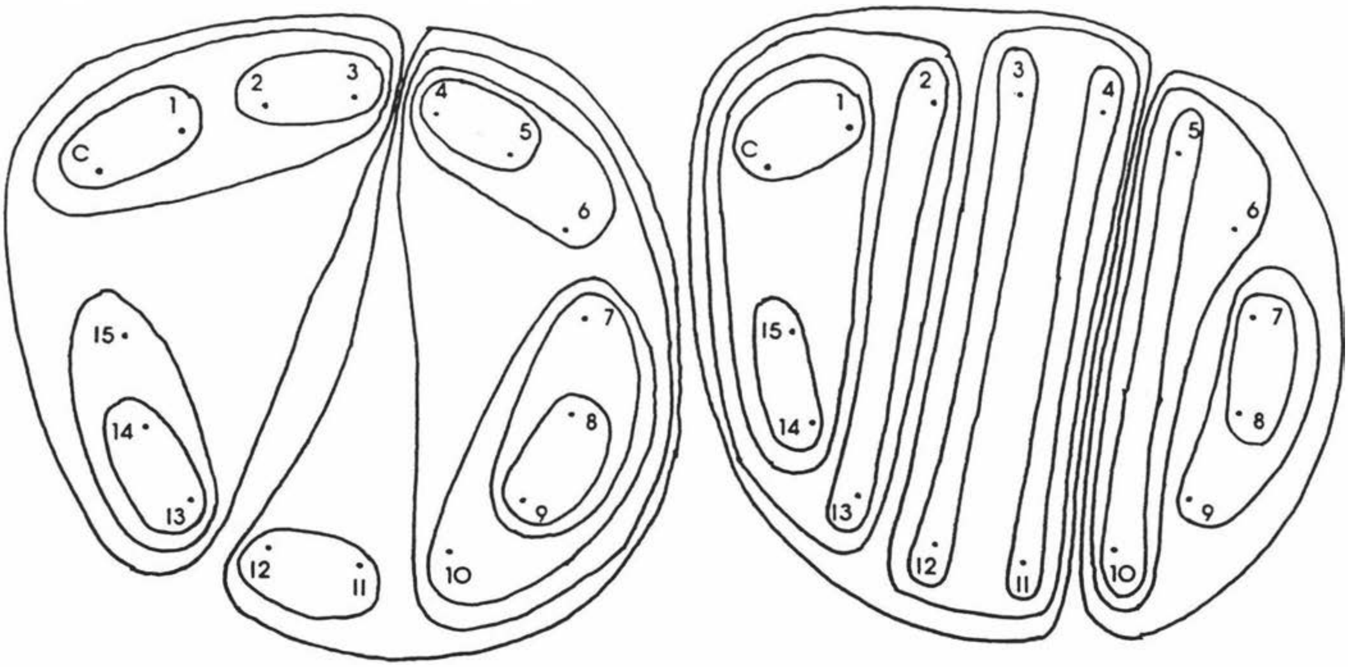
Michon [1972] compared the dendrograms for a set of complex sounds (audible radar signals), from subjects before and after they were tutored in recognising the sounds, and from the tutors. The dendrograms were not obtained directly, but by applying a clustering algorithm to dissimilarity matrices (themselves estimated by vote-counting triadic data): there is no guarantee that the subjects would have H-sorted the stimuli in the same way. Michon interpreted these derived hierarchies as search trees for identifying sounds – each branch being a test in the identification, "Is this feature present or not?" – and compared them on a single, pre-determined criterion, balance. Balanced trees, which involve fewest branches or tests as they are traversed from the root to the twigs, correspond to more efficient searches. As he expected, Michon found that such trees characterise expert listeners.

Several papers define "distance" functions (or spectra of functions) between dendrograms [Boorman & Olivier, 1973; Fowlkes & Mallows, 1983]. For dendrograms from M subjects, the result is an M-by-M table of distances, suitable for MDS. This approach has much in common with the first stage of the Messick-Tucker points-of-view analysis: the construction and scaling of a matrix of correlations between subjects' configurations. The perceptual space remains unexplored, and cannot contribute to the interpretation of differences between the hierarchies.

Coxon and Jones used this approach. They also performed in-depth analyses on the POOC H-sorts, examining each participant's sequence of merges in conjunction with a close semantic scrutiny of the interview transcript in order to determine the criteria or constructs used in each choice of groups to combine. This *tour-de-force* is time-consuming and does not generalise directly to other sets of elements.

I note now that there is no barrier to incorporating the INDSCAL model in the Method of Reconstructed Dyads. As in Chapter 4, it is only necessary to define a separate distance matrix D_m for each subject, defined in terms of the individual dimensional weights (4.1), to use in selecting the representative dyads $ij(I,J)$. Thus there are different coefficients $\delta_{m.g.ij.IJ}$ in (7.5). Also in (7.6), $D_{m.IJ.KL} = d_{m.ij(I.J)} - d_{m.kl\,(K.L)}$. Finally, the factors $(x_{ip} - x_{jp})/d_{ij}$ are replaced with $w_{mp}(x_{ip} - x_{jp})/d_{m.ij}$, as in (4.5), and the calculation of $\partial L_m / \partial w_{mp}$ proceeds as in (4.6).

Figure 7.7 Two aberrant merging sequences for D-15 stimuli, illustrating effect of dimensional weighting



Evidence for dimensional salience variations between subjects (elliptical rather than circular similarity contours) would be a tendency toward elongated groups during the intermediate stages of merging. Consider a situation where subject m must choose between combining piles I with J, or K with L, where $d_{ij(IJ)} = d_{kl(KL)}$, but $d_{m,ij}$ is reduced relative to $d_{m,kl}$ by a flattening of his or her perceptual space along some dimension, p; the likelihood increases of groups growing by accretion preferentially parallel to the p-th axis.

However, the effect is small, and its appearance relies on the right configuration. Recall the finding that recovery of individual variations is reduced in sorting and incomplete ranking data [Rao & Katz, 1971].

To be convinced that any w_{mp} values recovered are meaningful, warranting the additional degrees of freedom, we need more than an increased Likelihood for the configuration. There are two other desiderata:

(a) Non-arbitrary axes, i.e. loss of rotational symmetry. If the initial configuration is not aligned with the true axes, the process of maximising L should include a rotation to a better fit. If, alternatively, one can choose *any* set of dimensions and find subjects who arrange the stimuli into groups aligned roughly parallel with those axes, so that no set of axes is more appropriate than another, there is no justification for the INDSCAL model.

(b) We look for convergent validity: when the same subjects provide dissimilarity data in a different form, scaling them should produce the same dimensions and the same weights.

The POOC H-sorts satisfy (a). If rotated, the configuration finds its way back to the preferred axial orientation, albeit slowly. As for (b), within the available data pool there are regrettably no H-sorts and dissimilarity matrices from the same subjects.

Partly to test whether the dimensional weights obtained from H-sorts and from assessed dissimilarities were comparable, Barraclough elicited data from 18 subjects, for 9 colours. The nature and the analysis of the dissimilarity ratings – resulting in the familiar colour circle, figure 6.1 – were briefly discussed in Chapter 6.

I am not concerned with reconstructing the configuration from the H-sort data (which is just as well, since the small number of stimuli and their specific arrangement precluded doing so). Instead I analyse the two data sets together, so that the dissimilarity ratings hold X constant and L is maximised by altering the w_{mp} only. The INDSCAL model is used to approximate the more accurate IDIOSCAL model. Figure 7.8 shows the H-sort subject space. Subject points for nine subjects who contributed both forms of data are indicated by upper-case initials for the H-sorts, and lower-case initials for dissimilarities (the eight subjects provid ing dissimilarities only, and the one providing an H-sort only, are indicated by 'x' and 'X').

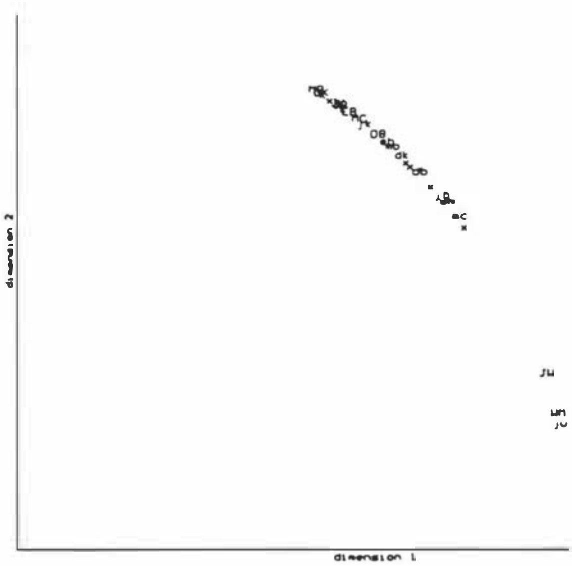
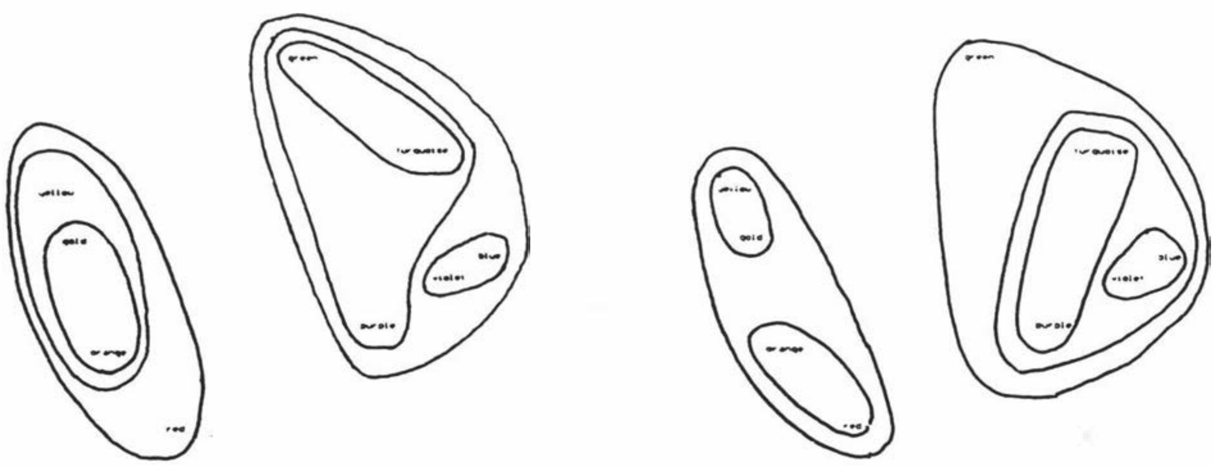


Figure 7.8 Subject space for nine colours

Two subjects, 'JW' and 'WM', had colour-vision deficiencies which showed up despite the saturated colours and the large angular extent of the stimuli (squares, 3° across, in both parts of the experiment). Otherwise, for those nine subjects overall, the correlation between w_{mp} as derived from their dissimilarity responses and from their H-sorts is only 0.674.

Figure 7.9 Aberrant merging sequences for subjects JW and WM, superimposed on the true configuration.



JM: ${}_4({}_2({}_1 5\ 6)^1\ 9)^2\ ({}_3 8\ 7)^3)^4\ ({}_7({}_6({}_5 3\ 2)^5\ 4)^6\ 1)^7$

WM: ${}_4({}_3({}_1 6\ 9)^1\ ({}_2 7\ 8)^2)^3\ 5)^4\ ({}_7({}_5 3\ 4)^5\ ({}_6 1\ 2)^6)^7$

I return at last to the D15 stimuli. H-sorting them is a rather imprecise means of obtaining individual differences. For reasons already discussed, only obvious colour deficiencies are likely to show up, in the aberrant H-sorts produced when the configuration collapses to a near-linear ellipse in the subject's colour space (the standard procedure for the D15 stimuli is no more likely to discriminate fine variations in colour perception; this is not its goal). Figure 4.9 was the result of applying the IDIOSCAL model to the D15 H-sorts, as well as to data elicited with the standard D15 procedure. Protanopes and deutanopes can be distinguished from judges with normal colour vision, and from each other.

Kinship

In this experiment, 11 subjects (4F, 7M) H-sorted a set of 15 kinship terms ("kincepts"), the targets of previous scaling studies [Burton & Nerlove, 1976; Burton & Romney, 1975; Fillenbaum & Rapoport, 1971; Jacobowitz, 1974; Lopes & Oden, 1980; Romney & D'Andrade, 1964; Rosenberg & Kim, 1975).

Table 7.4 15 kinship terms and their abbreviations.

gf	Grandfather	fa	father	un	uncle	br	brother
gm	Grandmother	mo	mother	au	aunt	sr	sister
gs	Grandson	sn	son	np	nephew	cs	cousin
gd	Granddaughter	dr	daughter	nc	niece		

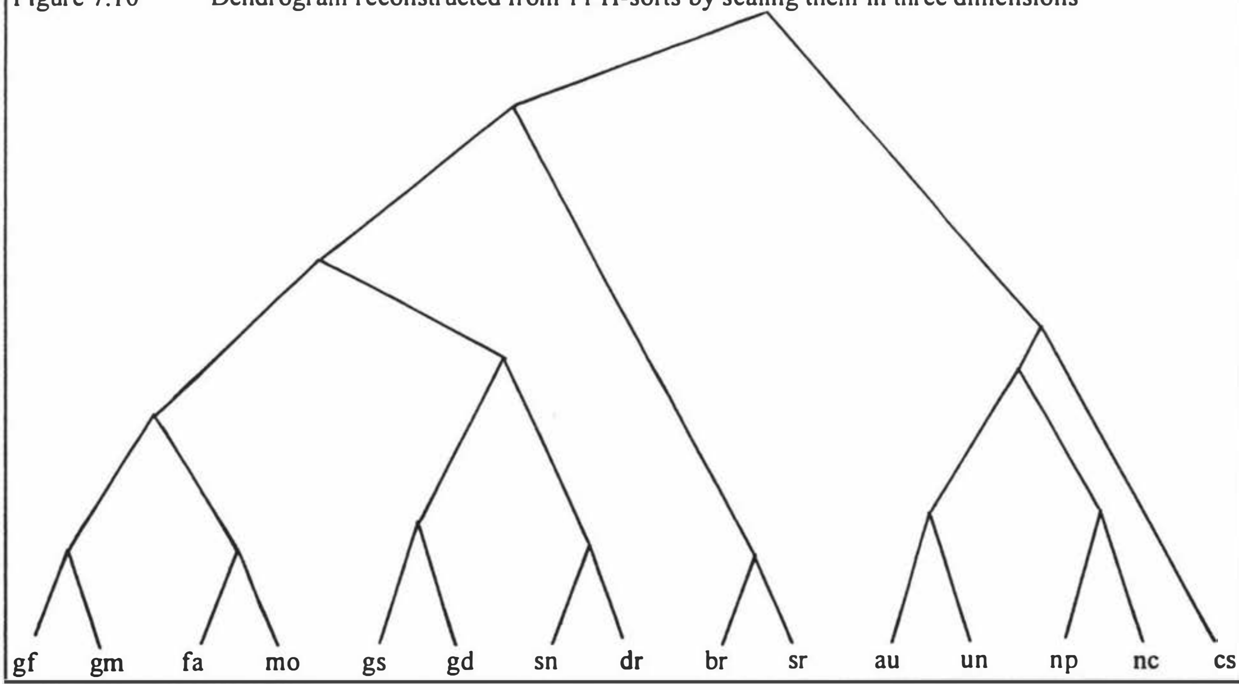
Table 7.5 Kinship H-sorts from 11 subjects.

$({}_{13}({}_{1}\text{br sr})^{1}\ ({}_{8}({}_{7}\text{np nc})^{7}\ \text{cs})^{8})^{13}\ ({}_{12}({}_{9}({}_{3}\text{dr sn})^{3}\ ({}_{6}\text{gs gd})^{6})^{9}\ ({}_{11}({}_{5}\text{un au})^{5}\ ({}_{10}({}_{2}\text{mo fa})^{2}\ ({}_{4}\text{gf gm})^{4})^{10})^{11})^{12}$
$({}_{13}({}_{12}({}_{7}\text{gs gd})^{7}\ ({}_{6}\text{gm gf})^{6})^{12}\ ({}_{11}({}_{5}\text{np nc})^{5}\ ({}_{10}\text{cs}({}_{4}\text{un au})^{4})^{10})^{11})^{13}\ ({}_{9}({}_{3}\text{br sr})^{3}\ ({}_{8}({}_{2}\text{fa sn})^{2}\ ({}_{1}\text{mo dr})^{1})^{8})^{9}$
$({}_{13}({}_{10}({}_{6}\text{gs gd})^{6}\ ({}_{3}\text{gm gf})^{3})^{10}\ ({}_{8}({}_{2}\text{mo fa})^{2}\ ({}_{4}\text{au un})^{4})^{8})^{13}\ ({}_{12}({}_{11}({}_{7}\text{np nc})^{7}\ ({}_{9}({}_{5}\text{br sr})^{5}\ ({}_{1}\text{sn dr})^{1})^{9})^{11}\ \text{cs})^{12}$
$({}_{11}({}_{1}\text{gm gf})^{1}\ ({}_{6}\text{fa mo})^{6})^{11}\ ({}_{13}({}_{10}({}_{4}\text{sn dr})^{4}\ ({}_{2}\text{gd gs})^{2})^{10}\ ({}_{12}({}_{8}({}_{7}\text{np nc})^{7}\ \text{cs})^{8}\ ({}_{9}({}_{5}\text{sr br})^{5}\ ({}_{3}\text{au un})^{3})^{9})^{12})^{13}$
$({}_{12}({}_{9}({}_{2}\text{gf fa})^{2}\ ({}_{3}\text{sn gs})^{3})^{9}\ ({}_{8}({}_{1}\text{gm mo})^{1}\ ({}_{4}\text{dr gd})^{4})^{8})^{12}\ ({}_{13}({}_{7}\text{br sr})^{7}\ ({}_{11}({}_{10}({}_{5}\text{un np})^{5}\ ({}_{6}\text{au nc})^{6})^{10}\ \text{cs})^{11})^{13}$
$({}_{13}({}_{5}({}_{3}\text{gm gf})^{3}\ ({}_{4}\text{gd gs})^{4})^{5}\ ({}_{9}({}_{8}({}_{6}\text{au un})^{6}\ ({}_{7}\text{nc np})^{7})^{8}\ \text{cs})^{9})^{13}\ ({}_{12}({}_{11}({}_{10}\text{dr sn})^{10}\ ({}_{1}\text{mo fa})^{1})^{11}\ ({}_{2}\text{br sr})^{2})^{12}$
$({}_{13}({}_{5}({}_{3}\text{gf gs})^{3}\ ({}_{4}\text{gm gd})^{4})^{5}\ ({}_{12}({}_{9}\text{un au})^{9}\ ({}_{2}\text{cs }({}_{1}\text{np nc})^{1})^{2})^{12})^{13}\ ({}_{11}({}_{10}({}_{8}\text{mo fa})^{8}\ ({}_{7}\text{sn br})^{7})^{10}\ ({}_{6}\text{dr sr})^{6})^{11}$
$({}_{13}({}_{7}\text{gs gd})^{7}\ ({}_{12}({}_{6}\text{br sr})^{6}\ ({}_{11}({}_{10}({}_{2}\text{fa gf})^{2}\ ({}_{1}\text{gm mo})^{1})^{10}\ ({}_{5}\text{dr sn})^{5})^{11})^{12})^{13}\ ({}_{9}\text{au }({}_{8}\text{un }({}_{4}\text{np}({}_{3}\text{nc cs})^{3})^{4})^{8})^{9}$
$({}_{12}({}_{3}\text{br sr})^{3}\ ({}_{11}({}_{5}\text{gd gs})^{5}\ ({}_{10}({}_{4}\text{sn dr})^{4}\ ({}_{9}({}_{1}\text{gm gf})^{1}\ ({}_{2}\text{fa mo})^{2})^{9})^{10})^{11})^{12}\ ({}_{13}({}_{7}\text{nc np})^{7}\ ({}_{8}\text{cs }({}_{6}\text{au un})^{6})^{8})^{13}$
$({}_{11}({}_{4}\text{fa mo})^{4}\ ({}_{10}({}_{3}\text{au un})^{3}\ ({}_{5}\text{gm gf})^{5})^{10})^{11}\ ({}_{12}({}_{2}\text{dr sn})^{2}\ ({}_{9}({}_{7}\text{nc np})^{7}\ ({}_{6}\text{gd gs})^{6})^{9})^{12}\ ({}_{8}\text{cs }({}_{1}\text{br sr})^{1})^{8}$
$({}_{13}({}_{10}({}_{5}\text{gs gd})^{5}\ ({}_{4}\text{gm gf})^{4})^{10}\ ({}_{12}({}_{3}\text{br sr})^{3}\ ({}_{11}({}_{1}\text{fa mo})^{1}\ ({}_{2}\text{dr sn})^{2})^{11})^{12})^{13}\ ({}_{9}({}_{8}\text{cs }({}_{6}\text{np nc})^{6})^{8}\ ({}_{7}\text{un an})^{7})^{9}$

These stimuli were printed in 36-pt lower-case Bookman, then pasted on strips of cardboard sized 3 cm by 13 cm. The instructions to subjects asked them to ignore these kincepts' various degrees of similarity to Self, and not to consider them as applied to their own family, or to any other families they knew, but to assess the similarity between various abstractions of "relative", free from context. More than any other study, this set of emotionally-loaded stimuli raised suspicions in subjects' minds as to the purpose of the test, and frequent reassurances were required that there were no "right" or "wrong" responses, nor any intention of tracking down latent psychopaths.

An aid to understanding multidimensional configurations is to portray them in the form of a dendrogram. Figure 7.10 was obtained by applying a hierarchical clustering algorithm to the values of d_{ij} reconstructed when $P = 3$ (assuming different dimensionality, or non-Euclidean space, would result in different distances). As with other forms of dissimilarity comparison data, there are no values of δ_{ij} to cluster directly. Like the aggregate dendrograms which Fillenbaum and Rapoport extracted from their data [1971, Chapter 4, figures 3 - 7], figure 7.10 is recognisable as the tree predicted by the Romney-D'Andrade componential analysis. See also Boorman and Olivier [1973] who consider kinship dendrograms at length to illustrate their distance functions for tree-comparison.

Figure 7.10 Dendrogram reconstructed from 11 H-sorts by scaling them in three dimensions



In agreement with the previous papers, the three dimensions are reciprocity, relative generation, and collaterality, in order of increasing salience. Carroll [1976] interprets the "relative generation" dimension in other ways (such as "dependence", since grandparents, children and grandchildren tend to be dependent on Self, economically and otherwise, more so than parents).

Many subjects began by pairing off the sexes – Mother with Father, Nephew with Niece, etc. Thus the distances between members of these pairs is small. When a fourth dimensions is added to the scaling, in the belief that it would become a Sex dimension, some separation of stimuli did take place – the male kincepts displaced one way and the females ones the other – but this was not clear-cut. Structure which had not fitted into other dimensions overflowed into this fourth one, obscuring the separation, so that identifying the dimension as 'Sex' required some prior knowledge.

Facial Expressions – Lightfoot series.

The stimuli in this experiment were the Lightfoot series of photographs of facial expressions. Original proof-sheets from the Kirkland archives were re-photographed and enlarged to passport-photograph size (unavoidably losing some details such as fine facial lines, which we expected would make identification of the expressions harder). Despite their antiquity, the 54 Lightfoot stimuli were an ideal test of the reliability of H-sort data and the method of analysis, since as well as the paper originally describing them [Engen, Levy & Schlosberg, 1958], subsets have been scaled by Abelson and Sermat [1962], Cliff and Young [1968], Gladstones [1962a, 1962b]. 54 being too large a set, I selected 24 photographs:

2, 4, 6, 7, 9, 13, 15, 16, 17, 18, 19, 20, 22, 24, 28, 29, 30, 31, 32, 36, 37, 42, 46, 55

This set maximises the overlap with the subsets used in the previous studies, to allow comparison between results.

24 was still too many for the limited patience of unpaid subjects (even the author had difficulty H-sorting 18). Our solution was to shuffle the photographs into two sub-sets of 12, to be H-sorted separately. The subsets were selected randomly afresh for each subject. For i, j in different sub-sets, $c_{m,ij}$ is treated as missing data. This does more than simplify the sorting task; it also ensures that each item appears in a variety of contexts, which ultimately has the effect that more of the dyads involving a given item undergo comparisons. What is good for

the "reconstructed dyad" analysis is bad for vote-counting. The varying contexts cause $c_{m,ij}$ to vary (i.e. the stage at which two items first co-occur is affected by the presence or absence of other items), leading to "noisy" e_{ij} values, obscuring the structure of the MDS solution.

Table 7.6 Two H-sorts of 12 photographs from each of 12 subjects, plus one H-sort of 18 photographs.

$(_{10}(_{2}28\ 29)^{2}(_{9}(_{8}(_{7}(_{6}32\ 15)^{6}55)^{7}(_{5}(_{4}(_{3}24\ 9)^{3}20)^{4}13)^{5})^{8}(_{1}46\ 31)^{1})^{9})^{10}\ 16$

$(_{6}(_{5}(_{4}(_{3}(_{2}(_{1}17\ 7)^{1}2)^{2}42)^{3}4)^{4}18)^{5}19)^{6}(_{10}(_{8}(_{7}36\ 37)^{7}20)^{8}(_{9}22\ 6)^{9})^{10}$

$(_{9}(_{2}(_{1}9\ 37)^{1}36)^{2}(_{7}17\ (_{6}7\ (_{3}28\ 24)^{3})^{6})^{7})^{9}(_{10}(_{8}18\ (_{4}55\ 42)^{4})^{8}(_{5}19\ 31)^{5})^{10}$

$(_{10}(_{5}2\ (_{1}6\ 29)^{1})^{5}(_{8}(_{7}(_{3}20\ 22)^{3}46)^{7}(_{2}30\ 13)^{2})^{8})^{10}(_{9}(_{6}4\ 15)^{6}(_{4}16\ 32)^{4})^{9}$

$(_{8}(_{6}16\ 32)^{6}(_{5}4\ (_{3}(_{1}15\ 18)^{1}55)^{3})^{5})^{8}(_{10}(_{4}17\ 46)^{4}(_{9}(_{7}31\ 6)^{7}(_{2}9\ 24)^{2})^{9})^{10}$

$(_{10}(_{9}(_{6}20\ (_{3}7\ 42)^{3})^{6}(_{7}13\ (_{5}30\ 37)^{5})^{7})^{9}(_{8}(_{1}22\ 36)^{1}(_{4}28\ (_{2}2\ 29)^{2})^{4})^{8})^{10}\ 16$

$(_{7}(_{6}20\ 19)^{6}(_{1}9\ 29)^{1})^{7}(_{10}(_{9}(_{2}42\ 4)^{2}(_{4}15\ 32)^{4})^{9}(_{8}(_{5}18\ 30)^{5}(_{3}2\ 31)^{3})^{8})^{10}$

$(_{10}(_{5}16\ (_{4}17\ 55)^{4})^{5}(_{9}(_{2}(_{1}7\ 36)^{1}24)^{2}(_{7}37\ (_{6}46\ 13)^{6})^{7})^{9})^{10}(_{8}(_{3}28\ 22)^{3}6)^{8}$

$(_{3}15\ 18)^{3}(_{10}(_{8}(_{1}30\ 37)^{1}13)^{8}(_{9}42\ (_{7}(_{5}20\ (_{4}9\ 36)^{4})^{5}(_{6}46\ (_{2}7\ 17)^{2})^{6})^{7})^{9})^{10}$

$(_{10}(_{8}22\ (_{1}6\ 29)^{1})^{8}(_{7}(_{2}2\ 31)^{2}(_{5}28\ 24)^{5})^{7})^{10}(_{9}(_{6}32\ 16)^{6}(_{4}4\ (^{3}19\ 55)^{3})^{4})^{9}$

$(_{6}(_{5}4\ 19)^{5}(_{1}15\ 18)^{1})^{6}(_{10}(_{2}37\ 42)^{2}(_{9}(_{3}30\ 9)^{3}(_{8}22\ (_{7}46\ (_{4}31\ 28)^{4})^{7})^{8})^{9})^{10}$

$(_{9}(_{4}(_{3}(_{2}13\ 24)^{2}17)^{3}2)^{4}(_{8}20\ 7)^{8})^{9}(_{7}36\ (_{1}6\ 29)^{1})^{7}(_{6}(_{5}55\ 16)^{5}32)^{6}$

$(_{10}(_{7}(_{3}2\ 20)^{3}(_{6}7\ 42)^{6})^{7}(_{8}(_{1}22\ 29)^{1}(_{2}6\ 28)^{2})^{8})^{10}(_{9}10\ (_{5}4\ (_{4}18\ 55)^{4})^{5})^{9}$

$(_{10}(_{9}(_{3}16\ 36)^{3}(_{5}17\ 37)^{5})^{9}(_{7}(_{2}13\ 30)^{2}(_{1}15\ 32)^{1})^{7})^{10}(_{8}(_{4}31\ 46)^{4}(_{6}9\ 24)^{6})^{8}$

$(_{9}(_{8}20\ (_{5}15\ (_{2}16\ 32)^{2})^{5})^{8}(_{7}7\ (_{6}13\ 31)^{6})^{7})^{9}(_{10}(_{3}22\ 9)^{3}(_{4}29\ (_{1}6\ 19)^{1})^{4})^{10}$

$(_{10}(_{9}24\ (_{3}36\ 37)^{3})^{9}(_{7}28\ (_{5}2\ 30)^{5})^{7})^{10}(_{8}42\ (_{6}(_{4}46\ 17)^{4}(_{2}4\ (_{1}55\ 18)^{1})^{2})^{6})^{8}$

$(_{10}(_{5}6\ 19)^{5}(_{9}(_{6}46\ 42)^{6}(_{4}13\ (_{1}9\ 7)^{1})^{4})^{9})^{10}(_{8}32\ (_{7}(_{3}4\ (_{2}15\ 18)^{2})^{3}55)^{7})^{8}$

$16\ (_{10}(_{8}(_{2}22\ 28)^{2}(_{5}(30\ (_{1}36\ 37)^{1})^{5})^{8}(_{9}(_{6}29\ (_{3}2\ 31)^{3})^{6}(_{7}20\ (_{4}17\ 24)^{4})^{7})^{9})^{10}$

$(_{10}(_{9}(_{7}42\ (_{2}4\ 55)^{2})^{7}(_{8}(_{1}2\ 31)^{1}(_{5}20\ 46)^{5})^{8})^{9}(_{6}29\ (_{4}7\ (_{3}22\ 28)^{3})^{4})^{6})^{10}\ 16$

$(_{9}(_{7}(_{6}13\ 17)^{6}(_{2}31\ (_{1}15\ 18)^{1})^{2})^{7}(_{5}30\ 37)^{5})^{9}(_{10}(_{8}36\ (_{4}9\ 24)^{4})^{8}(_{3}6\ 19)^{3})^{10}$

$(_{6}(_{5}9\ 37)^{5}20)^{6}(_{9}(_{8}22\ (_{7}24\ 31)^{7})^{8}6)^{9}(_{4}46\ (_{3}(_{2}(_{1}15\ 55)^{1}32)^{2}13)^{3})^{4}$

$(_{10}(_{6}7\ 36)^{6}(_{7}28\ 29)^{7})^{10}(_{9}(_{5}42\ (_{4}17\ 30)^{4})^{5}(_{8}2\ (_{3}16\ (_{2}(_{1}18\ 19)^{1}4)^{2})^{3})^{8})^{9}$

$6\ (_{9}(_{8}16\ (_{7}30\ (_{1}4\ 17)^{1})^{7})^{8}(_{4}42\ (_{2}19\ 55)^{2})^{4})^{9}(_{6}37\ (_{5}31\ (_{3}9\ 24)^{3})^{5})^{6}$

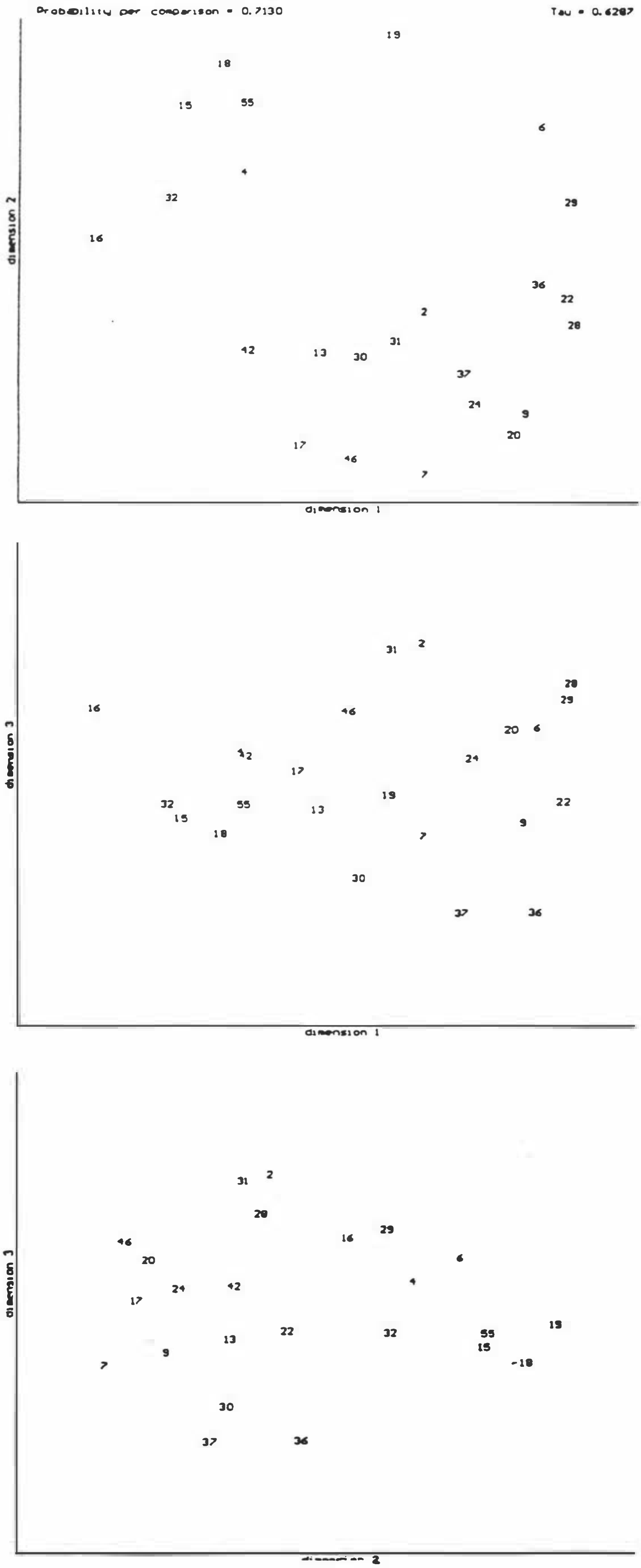
$(_{9}(_{3}(_{2}18\ 32)^{2}15)^{3}(_{6}(_{5}(_{4}7\ 46)^{4})13)^{5}36)^{6})^{9}(_{10}(_{8}20\ 22)^{8}(_{7}2\ (_{1}28\ 29)^{1})^{7})^{10}\ 15$

$(_{14}16\ (_{9}(_{7}32\ 18)^{7}55)^{9})^{14}\ldots$

$\ldots(_{13}(_{8}(_{1}31\ 2)^{1}29)^{8}(_{12}(_{11}(_{4}13\ 20)^{4}(_{6}(_{3}(_{2}46\ 17)^{2}7)^{3}9)^{6})^{11}(_{10}37\ (_{5}30\ 42)^{5})^{10})^{12})^{13}$

25 subjects undertook this double H-sort (13 for me, 12 for Barraclough). They were instructed to sort on the basis of similarity of *emotion*; and to ignore incidental features of the photographs, such as the varying sizes and contrasts of the prints, and the angles of the photographs. Stimuli 2 and 31 are different prints from the same negative (one slightly darker); I did not realise this when selecting the 24 stimuli, but the duplication has the advantage that the distances between x_2 and x_{31} give some indication of the experimental error

Figure 7.11 Three-dimensional configuration for 24 Lightfoot expressions (M = 51)



in the reconstructed configuration. 23 of the subjects also provided F-sorts for the complete set of 24 stimuli.

The results are in unexpectedly close agreement with Engen *et al.* Three dimensions for the solution were meaningful (figure 7.11). The effect of applying the INDSCAL model is that the configuration rotates (albeit slowly) to align these dimensions with the axes of the diagram.

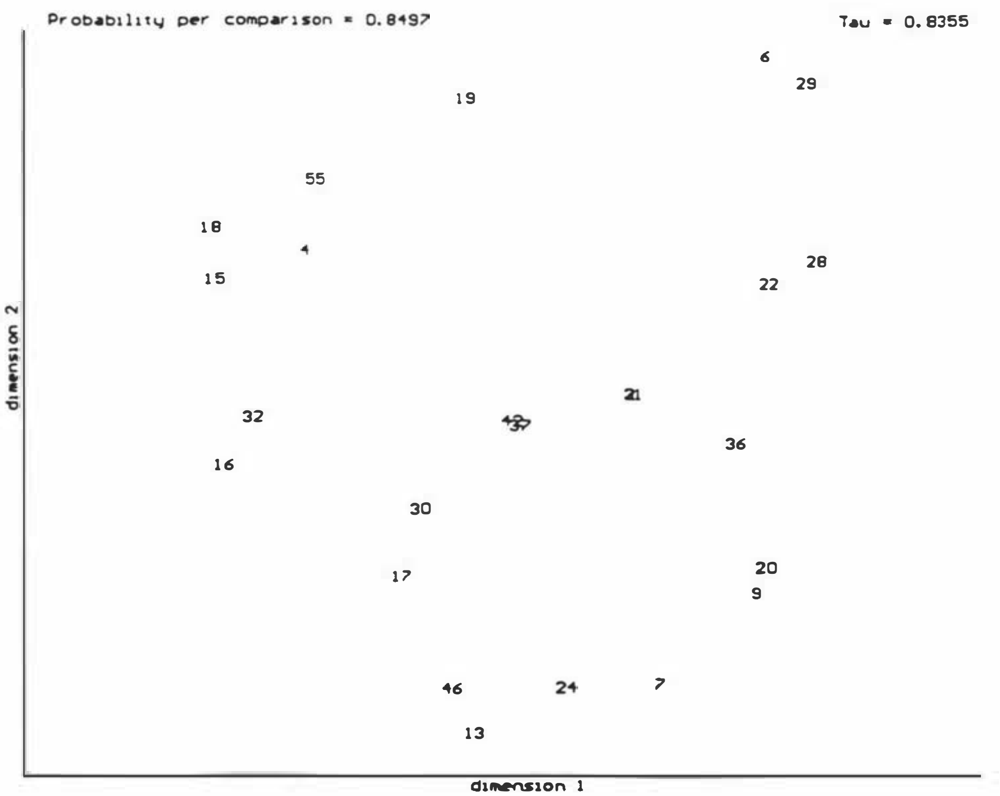
On first glance, the three-dimensional positions for the items agreed with triadic results for 10 items [Gladstones, 1962a, figures 3, 4]; closer examination showed high correlation with the three scales (Pleasantness-Unpleasantness, Attention-Rejection, Tension-Sleep) postulated by Engen *et al.* These scales are indicated in the figure (I incorporated the scale ratings in the configuration using the vector model and the Property-Fitting approach described in Chapter 6). The first dimension of the solution can be identified with the P-U scale. Following Osgood's [1966] analysis of cross-modal data (verbal identifications of 40 posed emotions), I tentatively identify the second as an Activation or Arousal or Intensity dimension.

The solution includes sequences of stimulus points with similar values on the Pleasantness and Activation axes, separated along the third axis: 37-30-42, 22-28, 4-32-16, 13-17-46, 9-20. The general impression from inspecting the corresponding expressions is that the distinguishing factor in these sequences is an increasing sense of spontaneity (Osgood suggests 'Control' for a third dimension, to discriminate, for instance, *loathing*, *rage*, and *horror*). This is not to rule out other labels. Gladstones suggests 'Expressionless-Mobile', and (rather despairingly), 'difficulty of interpretation', "[...] with expressions at one extreme which are difficult to interpret, and at the other extreme, stereotyped expressions which no one from a given culture would be likely to misunderstand" [Gladstones, 1962a, p.99]. If the object of the exercise had been to identify the dimensions beyond doubt, rather than a comparison of the "reconstructed dyads" analysis against earlier results derived from different forms of data, it would be easy to incorporate more stimuli in the configuration (it is a moot point whether the Lightfoot series is sufficiently representative of the range of normal, spontaneous facial expressiveness to make such an exercise worthwhile).

Clearly the A-R and S-T scales are not orthogonal (again, replicating Gladstones' results), but this does not justify conflating them to a single axis. It seems more reasonable to argue that both scales are linear combinations of two underlying dimensions. The matrix of dissimilarity ratings obtained for 13 Lightfoot photographs by Abelson and Sermat [1962] have generally been cited as two-dimensional [Borg & Lingoes, 1987; Shepard, 1962]. However, when I scaled those data in three dimensions, the positions on the third dimension for 11 items (the extent of the overlap between their stimulus set and the ones scaled in this work) showed a good degree of agreement with the H-sort analysis third dimension, the correlation being 0.784. This convergence suggests that the third dimension in both cases is genuine.
A final contribution to the identification of the axes comes from Cliff and Young [1968], who obtained ratings for "Intensity" for 13 Lightfoot photographs. Ten of the items they used also feature in this work's stimulus set, enough of an overlap to allow me to fit an "Intensity" vector to the configuration (again, using the joint scaling approach). It lay very nearly half-way between the vectors fitted to the A-R and S-T ratings from Engen *et al.*

A separate analysis of the F-sort data, using the Reconstructed Dyads method in a generalised form, discussed below, yields a similar solution: see figure 7.12.

Figure 7.12 Configuration for Lightfoot expressions derived from $M = 23$ F-sorts



I-Feel Faces.

Again, in this experiment the stimuli were photographs of facial expresssions, the faces being those of infants and young children. The 30 photographs come from the "I-FEEL" projective test distributed by the University of Colorado Health Services Center. When they are used as the test's designers intended, a subject describes the emotions displayed in each face; these descriptions are tabulated, and compared against base-line descriptions, to see if the subject is projecting undue amounts of happiness, sadness, anger, etc. onto the faces. Instead, Barraclough asked 20 subjects to H-sort them. For each subject, the stimuli were shuffled and split into two, each subset of 15 being H-sorted separately, resulting in 40 data sets. Barraclough also elicited F-sort data (using the complete set of 30 stimuli) from the same subjects.

I scaled the data in two and three dimensions. It is not certain whether young children actually experience three dimensions of emotion [Osgood, 1966], but this is not germane to the sorting, or indeed to the original I-Feel test, since what is at issue is the dimensionality of the emotions which subjects are projecting onto the stimuli... animal faces, or cars, or urban environments [Russell & Pratt, 1980] would presumably serve equally well.

A three-dimensional scaling, combining the two sets of data, leaves the first two dimensions more-or-less unchanged. At this point, the third dimension still awaits a satisfactory interpretation, and I will say no more about it.

There were no significant differences when I scaled the F-sort and H-sort data separately, and combinged, they produce the two-dimensional solution Figure 7.13(a). Comparison with figure 6.7 (derived independently, by scaling the I-Feel documentation's matrix of baseline emotion-attribution rates, treating it as cross-modal data) reveals a high level of agreement, apart from the overall curvature of the sorting configuration. The axes can be interpreted as a Pleasantness-Unpleasantness dimension, running horizontally from the joyful expressions of 101, 104, 106 and 124 across to 102, 117, 119, faces on the verge of tears or screams, if the photographs are any guide. There is a second Activation / Arousal dimension, running vertically from a sleeping child (122), up past the "Interest" label, to the surprised expressions

Figure 7.13(a). Configuration for the I-FEEL stimuli, combining H-sort and F-sort data ($M = 40$)

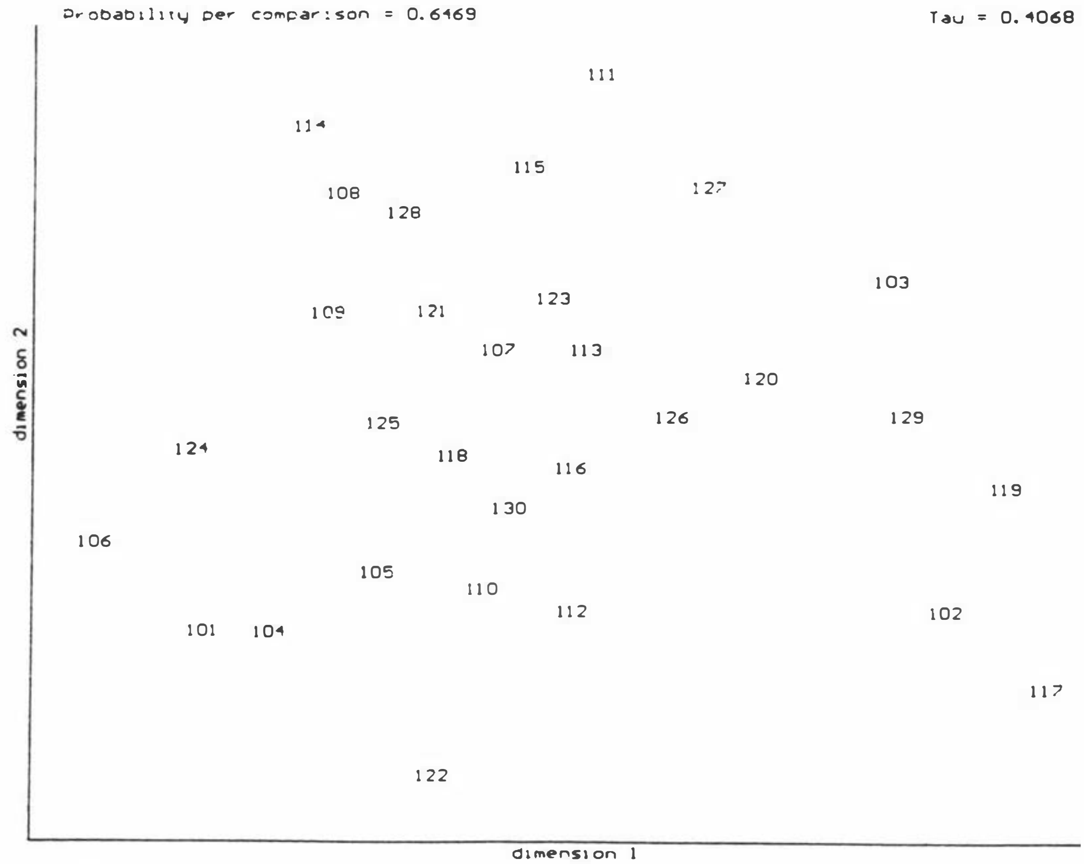
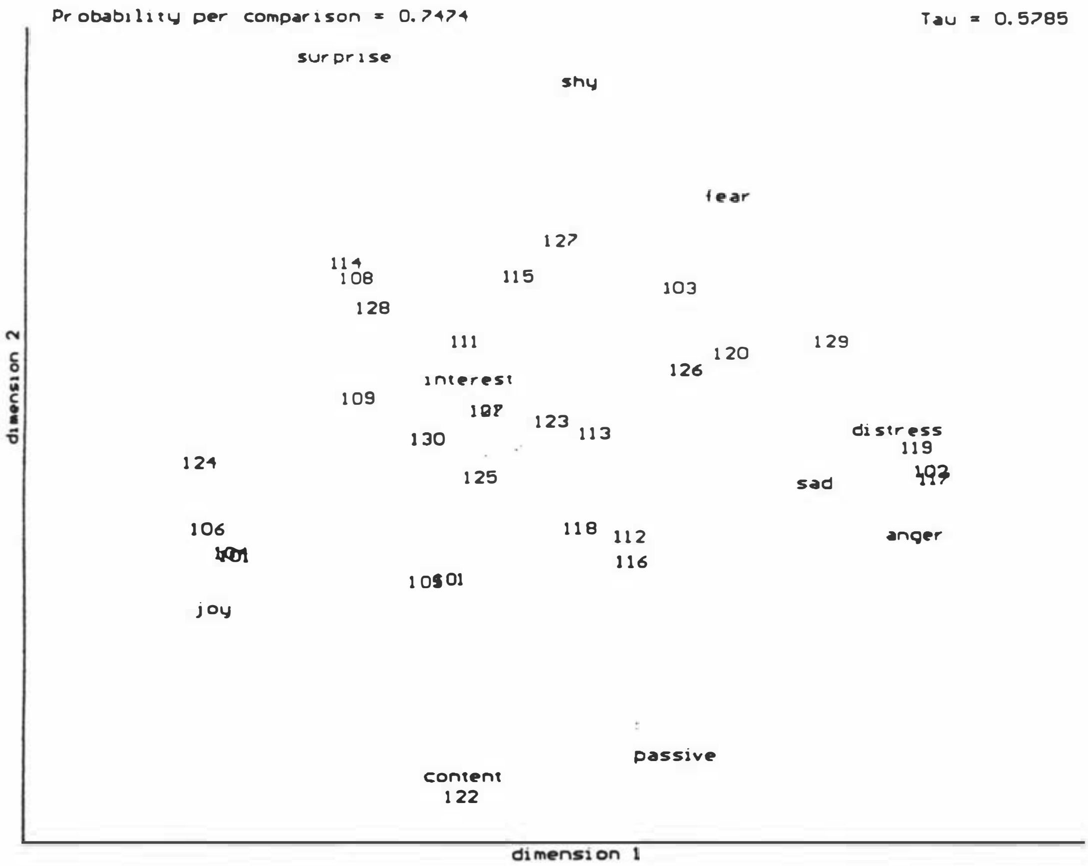


Figure 7.13 (b) I-FEEL stimuli, combining H-sorts and F-sorts with cross-modal emotion-attribution data



of 108 and 114. Given the compatibility of the sorting and H-sort responses with the emotion attribution data, I combined them, resulting in figure 7.13(c).

Woodblocks.

A jump from the most recent to the earliest H-sort data I processed. I am indebted to the 1994-1995 work of Barraclough for all of the data analysed in this section. Barraclough performed three series of H-sorts, using blocks of Paulownia wood, or photographs of blocks (as well as eliciting preference data for the same stimuli. A preliminary scaling of the preferences was shown in the previous chapter). The woodblocks, the photographs, and scale ratings for them were all provided by Bruce Glass of the Forest Research Institute.

Series 1 compared two versions of the hierarchical scaling procedure. To distinguish these, I call them *Synthetic* and *Analytic* H-sorting. The former has already been described. Groups are combined into larger groups in a process beginning with *N* groups of a single stimulus each and ends with a single group of *N* stimuli. The reverse process is Analytic H-sorting. A deck of *N* stimulus cards (or blocks) is presented to the subject, who is asked to separate them into 2 piles (of any size) comprising stimuli which "go together". Subsequent steps are to sub-divide either pile into two, continuing this until there are N single-stimulus piles.

I argue that this second version of H-sort can be treated in the same way as the first. At each step, the "distance" *between* piles should be greater than "distances" between stimuli *within* a given pile (distances in both cases being calculated with the help of an HCA). An economical use of the clustering algorithm is to view the stages of the successive sub-division in reverse time-order, i.e. as a Synthetic H-sort, calculating the Stress or Likelihood as before.

The earlier representations for H-sorts suffice to describe the Analytic version, so long as it is understood that time flows in the opposite direction, from the root of a dendrogram toward to twigs.

14 subjects sorted the 12 woodblocks with the Synthetic hierarchical procedure, and 14 with the Analytic procedure. The belief that the two procedures are equivalent was supported by

the results of processing the two data sets separately. The two configurations were very similar – closer to each other (and to the preference data solution) than either was to the configuration produced by vote-counting that data-set.

I therefore combined the data set for a single configuration, figure W.3 . In this the INDSCAL model is used. There is no evidence that the first and second 14 subjects (Synthetic and Analytic procedures) come from different populations: their respective spreads in subject space overlap. Figure W.2 illustrates two dimensions of the vote-counted result when the 28 data sets are combined. It is broadly similar but stimuli are clumped, obscuring much detail.

Barraclough finds the Analytic H-sort more convenient, and in light of its equivalence, has subsequently used it exclusively. This covers the data sets for the D15, the I-Feel, and the 9-colour elements. Some Dulux and half the Lightfoot data sets were obtained with the Analytic version; for the remainder I used the Synthetic version.

In a second series, Barraclough compared H-sorts of photographs of the woodblocks with the ones already obtained using the blocks themselves. 12 subjects sorted the photographs. Apart from the possibility of a lower average salience for the third dimension (colour), the configuration was not significantly different from figure W.3 (using "significant" in an informal sense). Photographs appear to be sorted in the same sequence as blocks.

A joint analysis of the data from these 12 subjects, pooled with the previous 28, plus the 40 preference rankings, was the source of figure W.5.

Glancing between this configuration, and the actual items, made it clear that the first dimension had something to do with grain spacing. At one extreme are blocks like 16 and 14 which are "quarter-sawn" so that the broad faces cut at right angles across growth rings, which show up as regular and most finely spaced. Spacing increases steadily along the axis to the widely-spaced growth rings of "plain-sawn" blocks (where they are parallel to the broad faces, revealing irregularity or "waviness").

The cluster (32, 36, 37) turn out, on inspection, to be marred by knots (marred in the assessment of most subjects, that is, but a minority preferred them to unknotted blocks). The second dimension, then, serves to separate blocks with knots from those without.

More formal identification of the dimensions is made possible by additional data in the form of ratings of the blocks, by 13 judges varying in their level of expertise, on a number of scales. I incorporate them in the joint analysis (averaging the ratings of 7 expert judges, and abandoning the scales which exhibit low correlations – the grain features they measure presumably contribute too little to dissimilarity in "wood space" to show up as dimensions). As discussed in Chapter 6, the result is a set of vectors, Y, aligned so that the projections of the stimuli onto the vector y_o for the o-th scale (i.e. the scalar products $x_i \cdot y_o$) have the maximal rank correlation with the corresponding scale ratings.

Figure W.5 was constructed using H-sorts, preferences, and scales. The axes are as follows:

* Dimension 1.	Grain angle and straightness; uniformity	
* Dimension 2.	Defects;	colour tones (black)
* Dimension 3.	Colour (shade);	colour tones (brown)

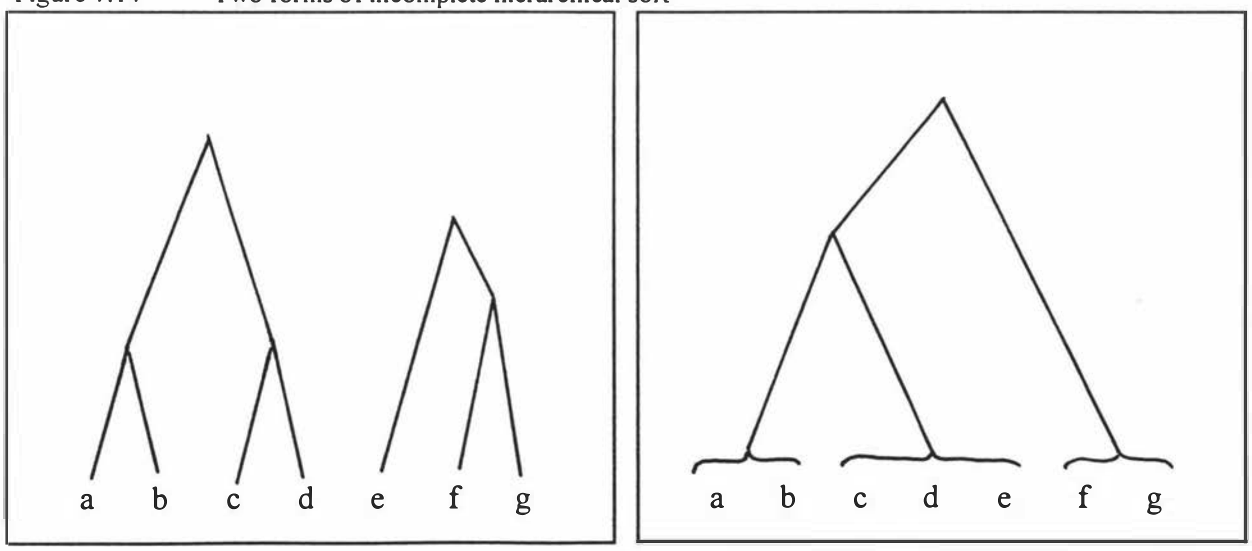
There were 12 additional items which had been rated on the scales but not involved in dissimilarity comparisons. Figure W.4 incorporates these items in the map: their positions are predicted purely on the basis of scale ratings. Barraclough set out to incorporate these extra items in wood space, adding to its rigidity, using the third series of H-sorts. These were conducted with photographs and Analytic H-sorting. Each H-sort used only 12 items (as with the expressions, 24 proved to be impractical). Some subjects sorted the additional set only, while for other subjects the old and new photographs together were divided randomly into two sets of 12, to be sorted separately. These data are scaled in figure W.5, verifying the predicted positions.

F-Sorting data.

When a subject sorts stimuli into groups, the algorithm he or she is following remains unknown and unobservable, but we are now in a position to propose a relatively tractable model for it. I agree with Takane [1980] that some kind of hypothesis about the process producing the sorts is a pre-requisite for calculating how likely it is that a given configuration produced them, and for constructing the best configuration.

Comparing procedures for large element sets, Rao and Katz [1971] consider a hybrid form of data (calling it "hierarchical subjective clustering"). This hints at the solution. Stoop [1986] asked 460 subjects to sort 18 social services into groups on the basis of similarity, then to merge those groups progressively until one group contained all the services. Sherman [1972] applied a similar procedure to adjectives. The result is an incomplete hierarchical sort, where the twigs and higher branches or the sorting tree (earlier stages of the merging sequence) are unknown. A final example of an incomplete hierarchical sort appears in Shweder [1972], where the subjects were 25 temple priests of Oriya, India, and the 81 stimuli they sorted were personality descriptors. A different form of incomplete H-sort, this time missing the later stages of merging (lower dendrogram branches) occurs when subjects are unwilling or unable to combine the last few groups, on the grounds of their extreme dissimilarity. There are cases of this in the Kinship and Lightfoot data.

Figure 7.14 Two forms of incomplete hierarchical sort



Without the subsequent mergings, Stoop's similar-service groups would be a slice through the dendrogram at a single height. This is my suggested interpretation for unconstrained sorting data: that they are incomplete hierarchical sorts. I envisage each subject grouping the stimuli according to an (unknown) hierarchical clustering algorithm, stopping at some stage to provide the experimenter with a snapshot of the merging sequence. The experimenter may elect to take the snapshot at a particular stage by specifying the number of groups in advance.

More complete forms of data are possible. For example, the subject's task might be to sort the stimuli into 13 groups; then to merge those groups until 8 remain; then to merge those 8 into 5; thereby capturing a series of slices through the subject's dendrograms. Alternatively, Clark [1968] asked subjects to sort stimuli (prepositions) into an unspecified number of groups, then to divide each group into sub-groups, thereby providing two snapshots of an Analytic hierarchical sort. Multiple sorts feature in Russell [1980], Russell and Bullock [1985, 1986], Ward [1977].

With groups growing by accretion or combination during the early, unobserved stages of a H-sort, there is no reason to expect them to be compact at the point it terminates. Stimuli *i* and *j* can belong to the same group, without being particularly similar, so long as there are stimuli located between them, providing a chain of high similarities. I am reminded of Wittgenstein's "family resemblance" account of how general concepts relate to the range of specific objects they encompass (he rejected the notion of a Platonic ideal which all specific instances must approximate).[2] Note also Vygotsky's concept of "proximal development".

This was the model Rao and Katz used [1971] used, simulating the sorting data for their comparison by interrupting a hierarchical clustering (Maximum HCA) of items.

---

[2] *"Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group but many common features drop out. When we next pass to ball-games, much that is common is retained, but much is lost. – Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. [...] We see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities in detail."* Ludwig Wittgenstein, *Philosphical Investigations.*

The treatment described so far for H-sorts has no difficulty accommodating incomplete data. If the exact sequence of group mergings which led up to a given partition of the stimuli is unknown, the HCA allows us to reconstruct it. For the sake of concreteness, suppose that subject m partitions the N stimuli into G(m) groups,

$$\wp(m) = \{ I_1, I_2, \dots I_{G(m)} \},$$

which can be expressed as a co-occurence matrix C_m:

$$c_{m,ij} = \begin{cases} 1 \text{ if } i,j \in I_k \text{ for some } I_k \in \wp(m) \\ 0 \text{ if } i \in I_k, j \in I_l \text{ for } k \neq l. \end{cases}$$

Interpreting this as distance comparisons in order to calculate $X^{(t+1)}$ depends on interpolating N - G(m) merging steps from $X^{(t)}$. Constraints are imposed on that interpolation by $\wp$. Each interpolated partition is $\wp_g = \{ I_{g,1}, I_{g,2}, \dots I_{g,N-g} \}$ where $0 \leq g < G(m)$, and the transition from $\wp_g$ to $\wp_{g+1}$ is interpolated by finding groups $I,J \in \wp_g$ such that d_{IJ} is minimal, and $c_{m,ij} = 1$ for the limiting dyad $ij(I,J)$ – i.e. the transition must be compatible with $\wp$. g such merging steps result in $\wp$. At each g, d_{IJ} should be less than d_{KL} for all pairs $K,L \in \wp_g$ which the data insist must go unmerged, i.e. $c_{m,kl} = 0$ for $kl(K,L)$. There is a contribution to Stress for any situations involving I,J,K,L, where $d_{KL} < d_{IJ}$ but K and L are prevented from merging by the data.

With the missing $\wp_g$ interpolated, specifying I, J, K, and L, Stress and Likelihood and their derivatives can be calculated as before. A clustering algorithm must be assumed: in what follows, the Link HCA is used. Analysis of sorting data can be expected to provide higher Likelihoods (and lower Stresses) than H-sort data, since in the former, I and J are chosen out of the available alternatives in order to minimise d_{IJ}, whereas in H-sorts they are provided by the subject, increasing the chances of encountering rank-order violations of the form $d_{KL} < d_{IJ}$.

Note that if a subject has provided no data at all – only the final stage of merging is observed, the single group of N stimuli – then the HCA proceeds to interpolate earlier stages, untrammelled by data, and produces a zero-Stress dendrogram in complete agreement with the configuration. We have already encountered this situation, in the treatment of H-sort data, when I considered the question of constructing an average dendrogram for comparison with outside data (e.g. kincepts).

More general forms of this model lead to the same result. The assumption that the subjects have followed a HCA is not essential: it is merely a way of stressing the similarity of the H-sort method. A much weaker assumption is that the subject starts with a threshold dissimilarity $Þ_m$, and proceeds to look for dyads (i,j) such that $\delta_{ij} < Þ_m$, merging the groups I and J to which i and j belong, not necessarily merging the closer dyads first. Let this continue until i and j are grouped together for all $\delta_{ij} < Þ_m$. The final composition of the groups, and the Stress, are not affected by the order of merging.

If one imagines $Þ_m$ starting small, so that each item is in a pile of its own, then progressively increased, this is an alternative model for the H-sort method; it results in the same series of slices through the dendrogram. However, these simpler models do *not* generalise to the Diameter HCA.

We have already seen results for the Lightfoot and I-Feel sets of expression stimuli, which displayed an encouraging resemblance to independent maps of expression space. Thanks to the generosity of other researchers, I can now apply this approach to several sets of previously-analysed data.

Occupations

In addition to the triads for 13 occupational titles, and the H-sorts and pairwise dissimilarities for 16, the POOC database includes sorting data from 31 subjects for an enlarged set of 32 titles (seven titles recur in all these sets, so they can be thought of as subsets of a single set of 39 stimuli). I applied MTRIAD to the 32-occupation sorting data. However, there is no outside basis for deciding whether the MTRIAD results are better or worse than vote-counting.

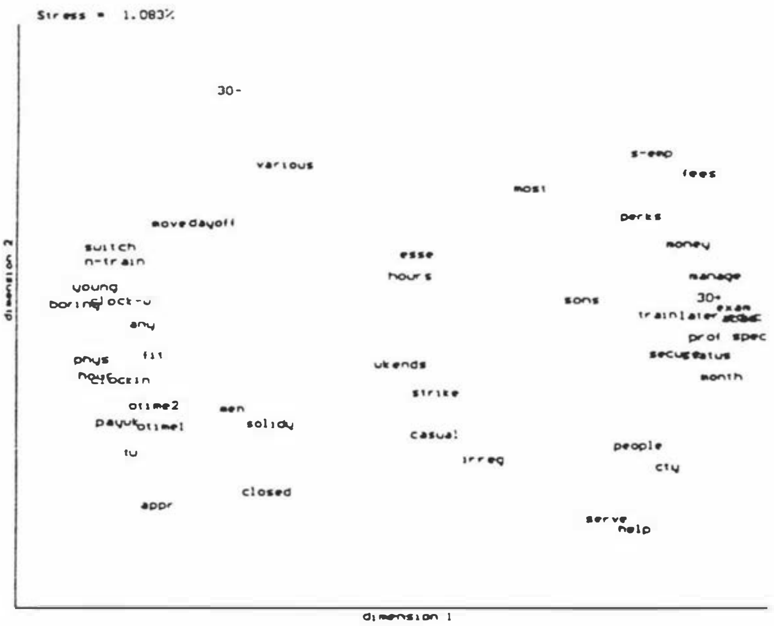
There is a second set of sorting data in the POOC files. 65 subjects sorted 50 predicates relating to occupations: simple statements such as “They work very long hours” and “They have a boring repetitive job” (these statements and abbreviations for them are listed in Table 7.7). The statements are general ones, to be grouped according to their similarities in their own right, independently of whatever specific jobs they might be applied to.

Table 7.7 Occupational predicate descriptions used in POOC sorting task (modified from Coxon and Jones [1978], Table 2.2, pp. 16-17)

	Abbreviations	Statements about occupations
1	strike	They would receive very little public support if they went on strike
2	hours	They work very long hours
3	manage	They are involved in managing people as part of their work
4	clock-w	They spend a lot of time at work clock-watching
5	s-emp	They are often self-employed
6	phys	They have mainly physical skills
7	serve	They provide a service to the community
8	closed	They have their job organised as a closed shop
9	appr	They have served their apprenticeship to become tradesmen
10	irreg	They have irregular hours
11	acad	They have to have a high standard of academic education
12	switch	They often switch their jobs
13	otime1	They earn a lot of their salary by working overtime
14	sons	They often encourage their sons to go into the same work as themselves
15	perks	They have a lot of fringe benefits and 'perks' in their job
16	most	Most people have thought of being one at some time in their lives
17	otime2	They get paid overtime for work they do out of normal hours
18	casual	They usually do their work dressed in ordinary casual clothes
19	tu	They have a strong trade union
20	dayoff	They often take the day off from work
21	any	Anyone with average intelligence could do the work for which they are paid
22	wkends	They often work at weekends
23	men	They are almost always men
24	paywk	They are paid by the week
25	clockin	They have to clock in and out of work with a time-card
26	30+	You expect them to be over 30 years old
27	prof	They regard themselves as professionals
28	train	They have to undertake a long arduous training for their job
29	help	They are involved in helping other people
30	boring	They have a boring repetitive job
31	month	They are paid by the month
32	money	They earn a great deal of money
33	status	They have a high social standing in the community
34	fees	They are not paid regularly, but earn fees for what they do
35	move	They often move into some other line of work after a few years
36	various	They have often had experience of working in various lines of work
37	cty	They tend to be active in the affairs of their local community
38	young	They earn a lot when young, but their incomes don't rise much after that
39	hour	They are paid by the hour
40	people	They build up relationships with other people as part of their work
41	spec	They work in a very specialised field
42	educ	They are required to have high educational qualifications
43	esse	Society could not continue to exist without them
44	n-train	No special training is required to be one
45	later	They do not earn much at first, but do have high incomes later on
46	30-	They are mostly younger than 30
47	fit	They have to be physically fit to do their job
48	secure	They have a secure job
49	exam	They have to pass difficult examinations
50	solidy	They have a tradition of solidarity with each other

Informants used more complex criteria to group these stimuli than in the case of the occupation titles. Coxon and Jones note [1978, p.53] that only 8% of subjects claimed to have based their groupings on one-dimensional rankings, as opposed to 31% for the occupation titles. By this consideration, the vote-count solution for the sorts is inadequate. Whether scaled in two dimensions (figure 7.15) or three, the descriptions are basically polarised along a single axis. There is a cluster of descriptions relating to high status and level of skill, and another of low status and skill. Closer examination reveals finer structure along the second dimension (Coxon & Jones, 1978, p. 52, figure 2.10), but the central gap along the first dimension remains unbridged.

Figure 7.15 Vote-count solution for occupational attributes F-sort



As it happens, independent evidence is available. Coxon and Jones [1978] conducted a pilot study using the 'sentence frame' method. 47 sociology students assessed how well each of the 50 statements described each of 20 occupations (these 20 being the 16 from the H-sort stimulus set, with four added). The assessment could be 'Always' (Statement *i* is always true of occupation *o*), 'Sometimes' and 'Never'.

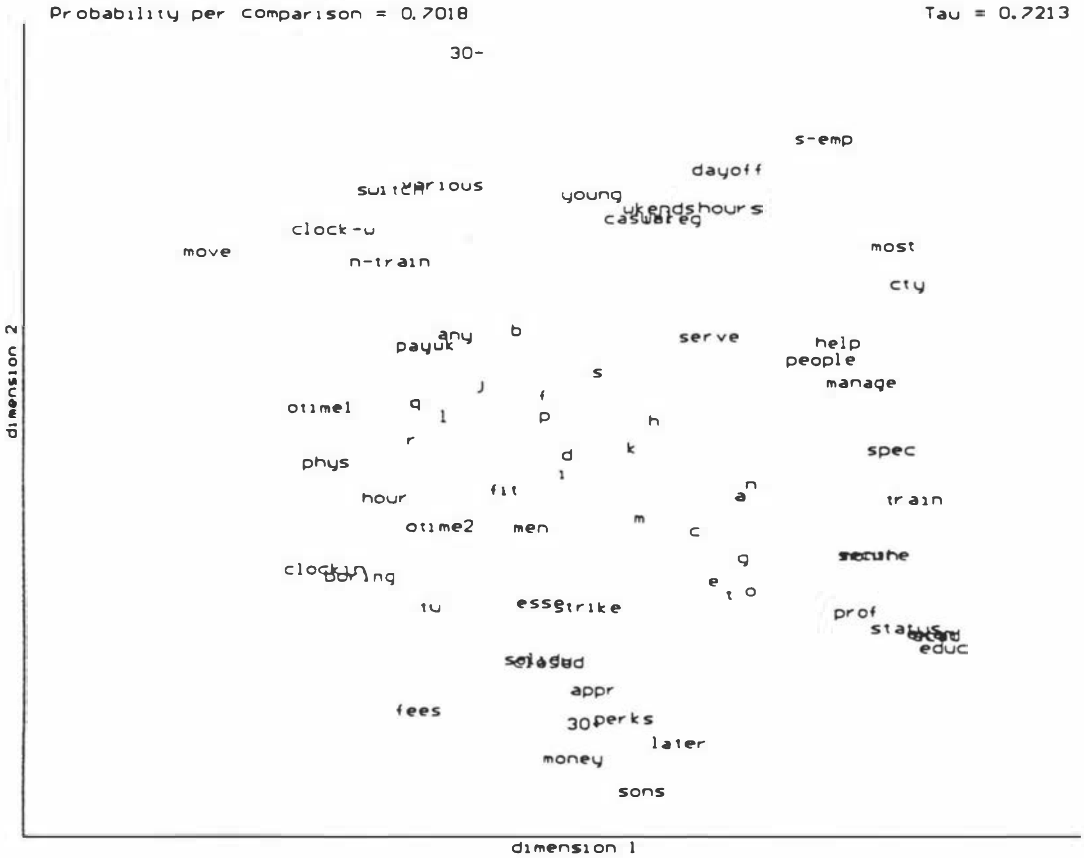
Table 5.2 of Coxon and Jones [1978, pp. 141-142] is a 50-by-20 matrix listing the percentage of students who answered 'always' for each description-title pair. These are cross-modal data, best treated as a sub-matrix of a 70-by-70 matrix of proximities. Given their incomplete nature, not much should be read into figure 7.16(a), the result of scaling these data in two dimensions. But both the overall distribution and the individual positions of the descriptions have much in common with figure 7.16(b), which comes from applying the reconstructed-dyad method to the sorting data.

For once, we have a situation where a circular arrangement of points around a central void is plausible, since using statements for stimuli restricts their distribution in 'occupation space'. They were selected to specify how a job can stand, how it differs from the average, so the central, neutral, non-committal region of the map should be sparsely occupied. To the extent that they focus on occupational qualities, the descriptions should form a circle.

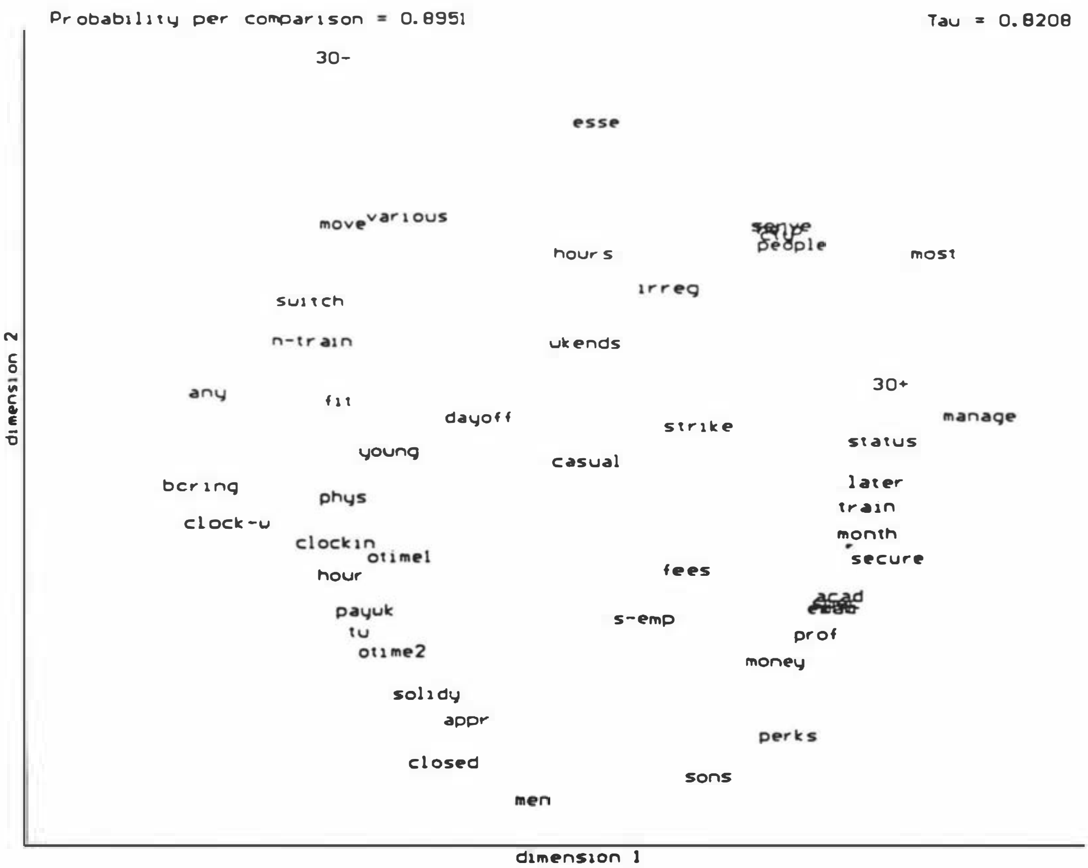
Figure 7.16(b) is a classic example of Guttman's circumplex, not only in its shape, but in the sequence of the descriptions around it. Any given nexus or combination of descriptions is linked to any other by a continuous, gradual shading of meaning, continuing in a complete circle. Any starting point is as good as another. Arbitrarily beginning in the upper right quadrant, one finds a small cluster of characteristics of what one might broadly call "social-work" jobs, e.g. 'They are involved in helping other people'. From there, through 'They build up relationships with other people as part of their job', one comes to managerial characteristics ('They are involved in managing people as part of their work'), then security and status ('They have a high social standing in the community'), which are characteristics shared with the 'professions'. The professions are singled out by the next descriptions – specialisation ('They work in a very specialised field') and training ('They have to undertake a long arduous training for their job'). This is quarter of the way around the circle from the starting point, and these characteristics are at right angles, as it were, to those originally encountered.

Following these linked, overlapping qualities further (in less detail), half way around the circle there are descriptions which emphasise physical fitness and physical skills instead of social ones. Still further, diametrically opposite the 'professional' descriptions, are ones pertaining to low status, non-specialisation ('Anyone with average intelligence could do the job for which they are paid'), and lack of training ('No special training is required to be one'). And so on, with opposite characteristics being diametrically opposite, back to the upper right quadrant.

Figure 7.16 Configurations for 50 occupational attributes: (a) using sentence-frame data



(b) F-sorts with $M = 65$



Power Strategies (data courtesy of Willem van der Kloot).

The stimuli in this study were 16 strategies, described in Dutch, for getting one's own way. The English translations are (1) to manipulate, (2) to hint, (3) to put someone in a good mood, (4) to deceive, (5) to look sincere, (6) to evade, (7) to threaten, (8) to pose a *fait accompli*, (9) to assert, (10) to persist, (11) to state simply, (12) to claim expertise, (13) to reason, (14) to compromise, (15) to bargain, and (16) to persuade. These were sorted by 25 subjects [van der Kloot & van Herk, 1991].

A two-dimensional solution is the result, figure 7.17(a). Two polarities stand out in the configuration, between Cooperative strategies ('compromise', 'bargain') and Coercive ones ('threaten'), and between Overt strategies ('state', 'reason') and Covert, Machiavellian ones ('evade', 'look sincere'). These provide possible interpretations for the axes.

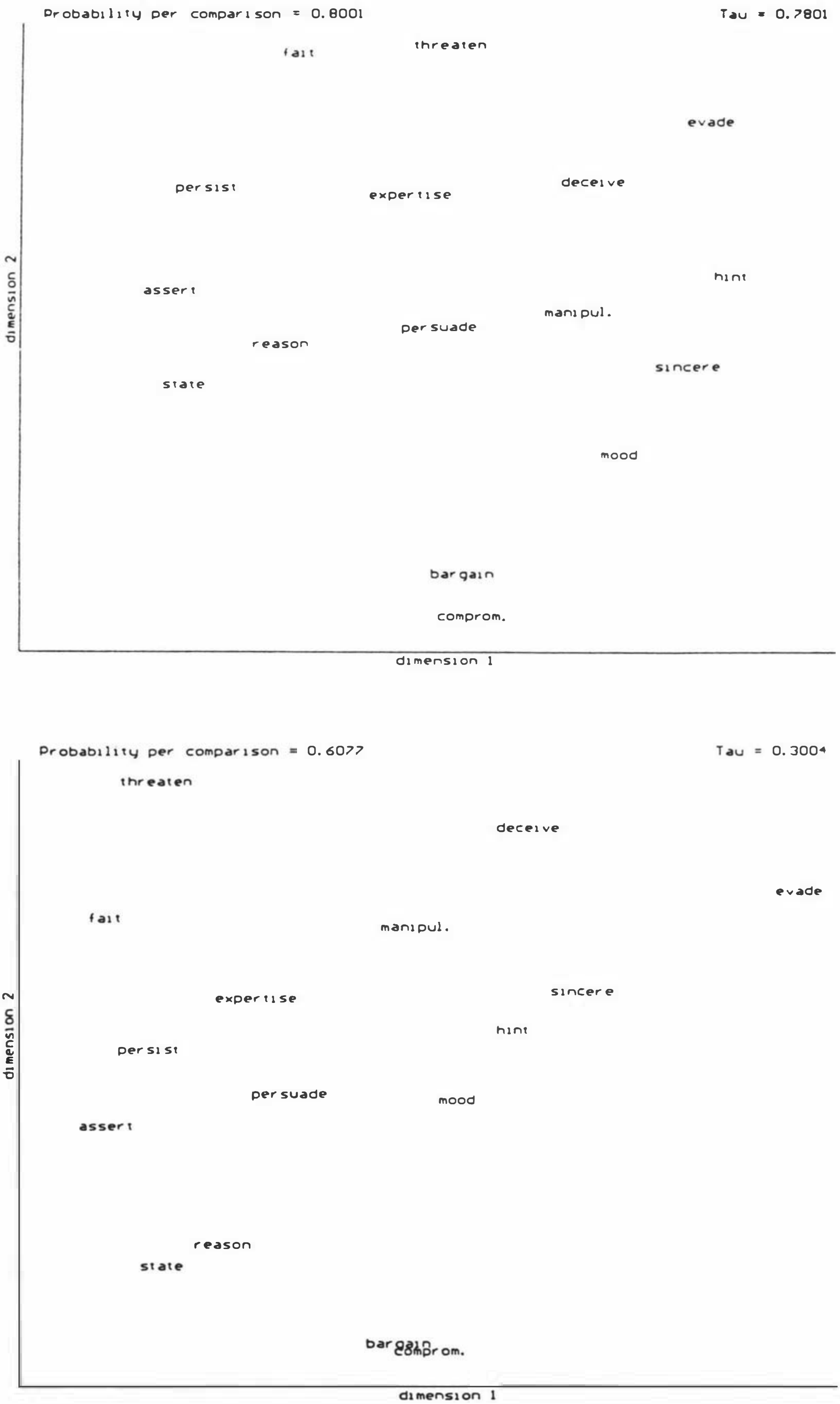
Van der Kloot and van Herk also collected pairwise dissimilarity ratings from the same subjects for the same stimuli, providing an opportunity for a direct comparison between solutions. The configuration for the dissimilarities is not a circumplex: figure 7.17(b) (see also their figure 8, p. 576). Like figure 7.17(a), it includes stimuli in the central region.

Kinship

Rosenberg and Kim [1975] elicited sorting data from 340 subjects for the 15 kinship terms already considered. Three of the anticipated four dimensions are present in their analysis, but fine gradations are lost. Their solution consists basically of an equilateral triangle of three clusters (nuclear family, grandparents and grandchildren, and the collateral relatives: Aunt, Nephew, Cousin, etc.), further subdivided on the Sex dimension, to form six clusters located at the corners of a triangular prism. Thus there is a "relative generation" dimension, which fails however to separate Brother and Sister (same generation as Self) from parents and children, or Cousin from the other collaterals. Fine gradations are missing.

This appears to be an artifact of their analytical method rather than a feature of the data. The groups created by 85 subjects (the female, single-sort fraction of the total subject population)

Figure 7.17 Configurations for 16 power strategies, using F-sorts (a, top) and dissimilarities (b, bottom)



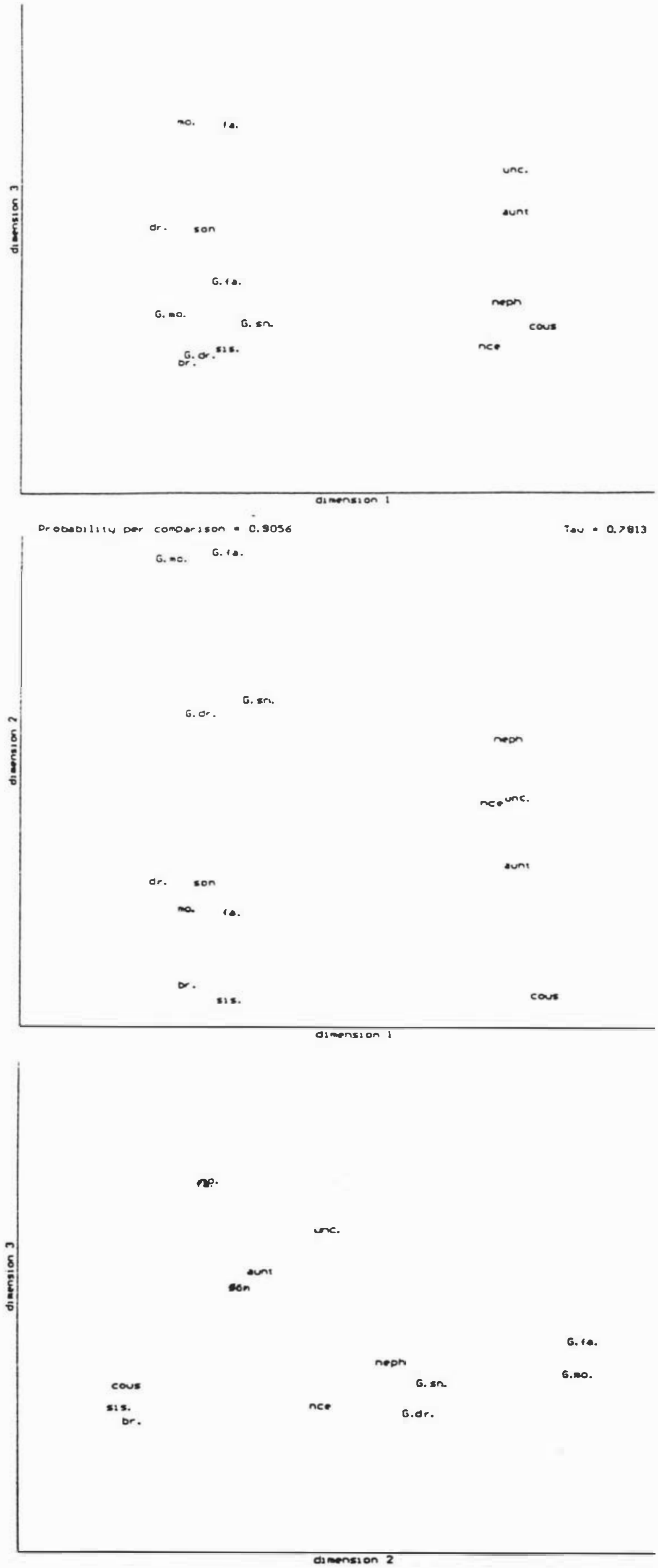
were published [Rosenberg, 1982, Table 7.1, pp. 121-123]. Presumably this is the same list of 85 sortings which Carroll [1976] analysed using non-spatial and hybrid models. The published list contains a number of respondents who separated siblings from parents and children, and Cousin from other collaterals, and others who distinguished between three levels of collaterality (direct ancestors / descendants; siblings; and Cousin, Aunt, Nephew, etc). However, these distinctions are only discernible in a subject-by-subject inspection of the groups: they are lost in the summed co-occurence matrix.

Since Rosenberg *et al* had not found the sub-population of 85 F-sorts to be significantly different from the other 255, I subjected them to MTRIAD. Figure 7.18 shows the results. Finer gradations become apparent when the context of each subject's distinctions is thus preserved. Note that three dimensions are interpretable; the data contain "absolute generation" distinctions (subjects who distinguished parents from children, grandparents from grandchildren, Aunt from Niece), as well as the relative generation dimension, although vote-counting obscures the former. Vote-counting also exaggerates the salience of the Sex dimension: of the 85 subjects, only five sorted the kincepts into a male and a female group (with Cousin separate). Another four partially sorted by Sex.

Alternative models for sorting data: Scales

The Rosenberg and Kim kincept data are not typical sorts. I suspect, after examining the data, that when the dimensions are as apparent and the stimuli spaced along them as regularly as in this case, then subjects may group stimuli together which share a value on some axis of their internal representations, rather than group them according to proximities. This is an alternative model for sorting data. If it applies, then essentially the groups are scale ratings (though in unknown order), suitable for processing with the stronger assumptions of Homogeneity analysis. Parallel hyperplanar slices through the configuration should separate each subject's groups. Coxon and Jones were perhaps thinking along these lines when they prompted some of their sorting-task informants to arrange the piles of stimuli, having sorted them, in some kind of order.

Figure 7.18 Three-dimensional solution to 85 F-sorts of kinship terms (Rosenberg *et al*)



Alternative Models: Pick Any/*N*

During the 1970s, a number of studies of personality and trait attribution were conducted in which subjects grouped trait-descriptive words and phrases together so that the descriptors in a given pile were similar in the sense of all applying to a particular person (real or imagined). "Sorting by exemplar", this might be called. The distances between descriptor points play no part in the formation of the groups or the reconstruction of a configuration of them. Thus, the data should not be affected by the addition of a new descriptor to the stimulus set; there are no chains of linked stimuli, no chance for a newly-provided **stimulus** to combine groups by bridging the gap between them (these properties apply to the previous model as well).

In some of these studies [Rosenberg & Sedlak, 1972; Wing & Nelson, 1972], descriptors could appear in several groups, or in none. Such sets of data are not strictly sorting data, in the sense of exhaustive, exclusive partitions; they are covered by the pick any/*N* variant of the hub-and-spoke model. In Chapter 6 I interpreted such data as low-resolution proximities or vector products: the *i*-th descriptor either belongs to the *o*-th group, i.e. x_i is close to the corresponding personality "ideal point" x_o, or it is less close, and doesn't belong (alternatively, it lies far enough / not far enough out along the corresponding vector). d_{io} is compared with some constant þ.

Other studies required that each descriptor be applied to one and only one personality [Friendly & Glucksberg, 1970; Rosenberg, Nelson & Vivekananthan, 1968, Rosenberg & Olshan]. It is less clear how to model the scaling process in this case (though one might start by arguing that it is still a hub-and-spoke situation, with the requirement of exclusiveness dividing "attribute space" into a Voronoi tessellation, the *i*-th descriptor belonging to the *o*-th group if and only if $d_{io} < d_{iq}$ for all other ideal points x_q).

This seems an appropriate place to mention Takane's program for the scaling of sorting data, MDSORT [Takane, 1980, 1981a]. The underlying model is not spelled out, but the program works by minimising the distances, in each subject's sorting, from the centroid of each group to its constituent stimuli, i.e. by arranging points to maximise group compactness, summed over subjects. Takane observes that this can be considered as a special case of Coomb's unfolding model. In effect this is a pick any/N analysis, treating group centroids as the ideal points.

Alternative models: Stochastic

A variant of the sorting procedure presents stimuli to the subject sequentially, in a randomised order. For each stimulus, the subject has the choice of starting a new group with it as the first member, or adding it to an existing pile, if it is close enough to any of them (only the former option exists when the first stimulus is presented). Hollins, Faldowski, Rao & Young [1993] applied this variant procedure to tactile textures.

Another example, arguably, is Struhsaker's observations of vervet monkey sleeping groups (published in Boorman & Arabie, 1973]. One imagines the troupe of monkeys, one or two dozen of them, preparing to settle down for the night. The first to feel sleepy chooses itself a tree. One by one, each of the remaining monkeys must decide which sleeping group it prefers to join, or if it prefers none of them, it can initiate a new sleeping group in a tree of its own.

Suppose we model the monkeys' preferences in terms of social distance. The distance between monkey i and sleeping group J could be defined as the minimum of the distances between i and the members of J:

$$d_{iJ} = \min_{j \in J} d_{ij} \qquad \text{or alternatively,} \qquad d_{iJ} = \max_{j \in J} d_{ij}$$

when the monkey's social distance is limited by the member of J it dislikes least. Write this as d_{IJ}, I being the group containing only i. In either case, there is some threshold value þ: if $d_{IJ} <$ þ, for at least one J, then I merges with the group J for which d_{IJ} is minimal; otherwise, it remains separate (and the monkey starts a new sleeping group).

These definitions of inter-group distances make it clear that this model of group formation has much in common with the interrupted hierarchical clustering model. However, groups can only merge if one or both are single items. Once created, two separate groups do not merge upon pres-entation of a new stimulus close to them both. For a constant distance matrix D, the final partition depends on the order of stimulus presentation: whether i and j are grouped together depends on whether other stimuli were presented in the right order to build a bridge between them.

Unfortunately this model cannot be applied to the two available data sets, Struhsaker's and that of Hollins *et al*, because both consist only of the final partition of the elements into groups, not the sequence in which elements were presented. Knowing that sequence would remove much of the rôle of guesswork and reconstructed distances in reducing the sorts to the level of dyad comparisons. Instead, I resort to the interrupted hierarchical model, despite the differences pointed out in the previous paragraph, hoping that the effects of presentation order introduce only additional error, and not systematic distortions.

In Hollins *et al*, 19 subjects followed this procedure to sort 17 tactile texture (bark, tile, etc). The published configuration, derived from the co-occurence matrix, is three-dimensional, but two tight clusters account for the majority of the stimuli; the rest of the configuration is only sparsely occupied (figure 7.19(a)). Subjects also rated the textures on five scales. Vector-model scaling of these ratings on their own also indicates three dimensions, but with less bunching of the elements: see figure 7.19(b). When the subject's sorts are analysed with MTRIAD (I am grateful to Faldowski and Hollins for access to their data), stimuli are spread more evenly through the three-dimensional result, figure 7.19(c).

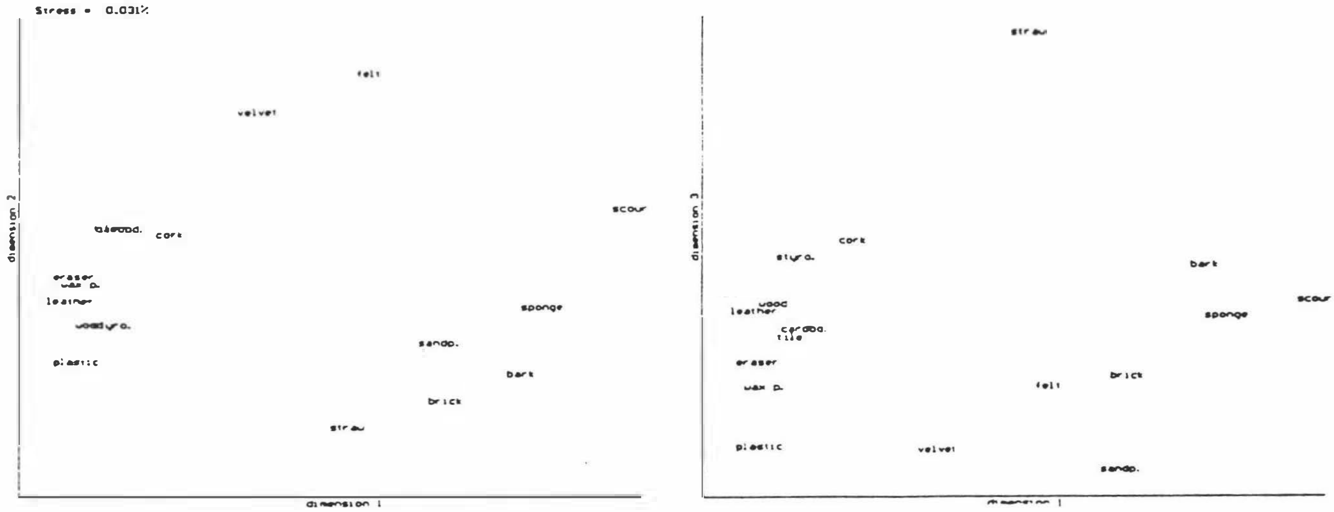
Alternative models: Low-resolution Proximities

This model interprets the sorts as low-resolution proximity data: the co-occurence matrix for each subject is assumed to be a table of dissimilarities,

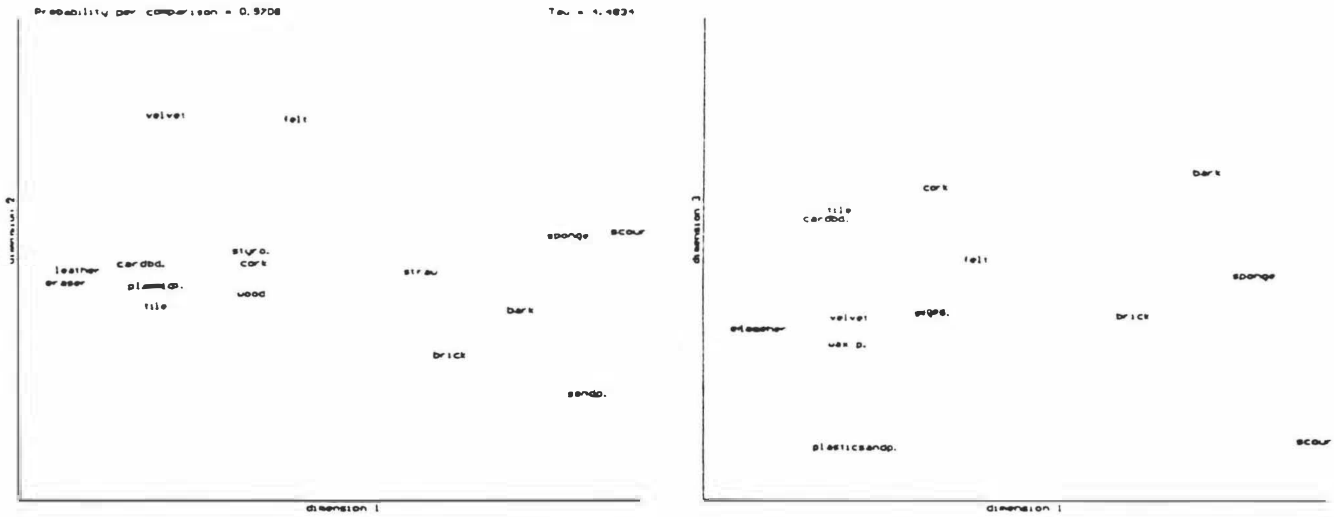
$$c_{m.ij} = \begin{cases} 1 \text{ if } \delta_{m.ij} < þ_m \\ 0 \text{ otherwise} \end{cases}$$

for some cut-off dissimilarity $þ_m$, which varies between subjects. In other words, $c_{m.ij} > c_{m.kl}$ is interpreted as a comparison between dyads, $(i,j) \ll (k,l)$. The matrices C_m can be analysed by INDSCAL to preserve dimensional weights [Rao & Katz, 1971], but are more usually averaged to form a matrix E. This model is taken for granted by every study of sorting data in which MDS techniques are applied to E. Hojo [1993] examined it explicitly and described a

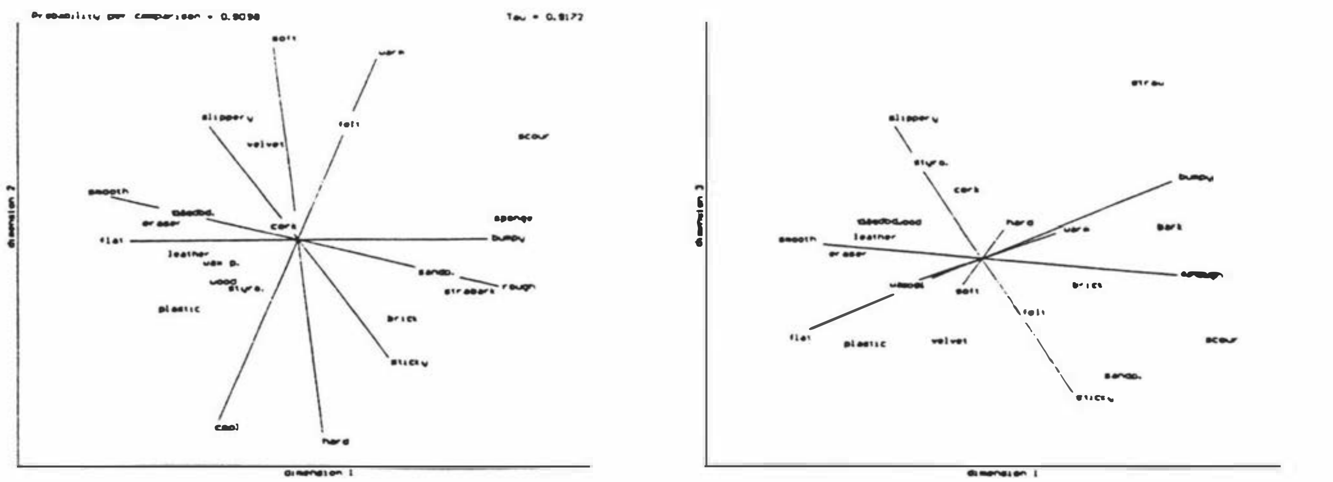
Figure 7.19 Configurations for 17 tactile stimuli. (a) Vote-count solution



(b) scale-based solution (vector model)



(c) reconstructed-dyad solution



Stress-minimising technique: contributions to Stress come from clashes between reconstructed distances and the data of the form $d_{ij} < d_{kl}$ despite $c_{m,ij} > c_{m,kl}$. Not surprisingly, Hojo's solutions are very close to those produced by standard treatment of his data.

A commonly-cited precedent for using a co-occurence matrix as a substitute for pairwise comparisons is Ward [1977], in which the two forms of data were compared and their configurations found to be reasonably similar. Regrettably, the exact degree of similarity remains in doubt, since the paper does not display the configurations or put a figure to the correlation between them. Ward's data is no longer available; we cannot tell whether other forms of analysis would have worked better than applying MDS to E.

Rosenberg [1982] reviewed the comparisons between these kinds of data. As support for sorting co-occurences as a substitute for pairwise comparison, he cites unpublished masterate theses by Drasgow and Davison, and the spatial-sorting work of Bricker and Pruzansky [1970].

Conversely, Sherman [1972] followed Bricker and Pruzansky by eliciting pairwise comparisons and partial hierarchies for a set of 20 trait-descriptive words. They differed dramatically. The configuration for the latter was degenerate, stimuli collapsing into one of two tight clusters.

Paddick [1978] used the sorting method on 18 odour stimuli. The 3 subjects who sorted the odours had already rated them on a number of Semantic Differential scales, but despite their familiarity with the stimuli, there was little agreement between the three-dimensional configurations derived from the sorts and the Semantic Differentials. Only the first dimension of the former lent itself to interpretation.

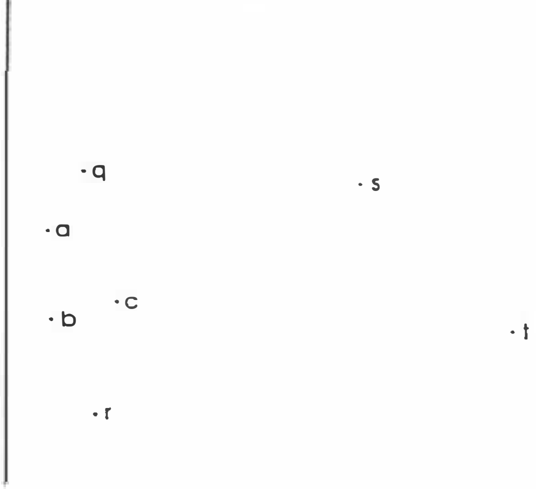
At the risk of boring the reader, I now recapitulate points already made about the flaws in the low-resolution proximity model.

I have already noted that grouping i and j together when $\delta_{ij} < þ_m$ results in the same partition that would be produced by applying the Link HCA, merging groups in sequence, closest first.

In other words, if $c_{m,ij} = 1$, then $h_{ij} < \text{þ}_m$, where h_{ij} is a 'hierarchical clustering metric' as defined by Johnson [1967] (Johnson uses the symbol $d(i,j)$, but that would invite confusion). This does not constrain δ_{ij} directly. A group I may include $\delta_{ij} > \text{þ}_m$ (for $|I| > 2$) because shared membership is transitive: if $\delta_{ik} < \text{þ}_m$, $\delta_{jk} < \text{þ}_m$, then i and k are grouped together, as are j and k, so i and j are necessarily in the same group whatever the value of δ_{ij}. I venture the statement that the larger $|I|$, the more likely it is that the mean dissimilarity between members of I exceeds þ_m; in fact, that on average, the mean dissimilarity between members of I increases with $|I|$.

In the illustration, the presence in stimulus space of elements a, b, etc increases the probability of q and r ending up in the same group, reducing the apparent e_{qr} in comparison with e_{st}. This effect is not cancelled out by aggregating data from subjects using different þ_m, nor is there reason to expect other forms of individual variation to eliminate it.

Figure 7.20 Hypothetical configuration



The second problem springs from the paucity of low-similarity information in sorting data: sufficiently dissimilar elements i and j might only be grouped together by one or two subjects. However, MDS on E does not allow for the larger statistical fluctuations when e_{ij} is small, and treats our ignorance about d_{ij} as spurious constraints. It is precisely those low values of e_{ij} in which we can have least confidence which MDS depends on the most. For examples of this second artifact, if more are needed, see the behaviours scaled in Burton [1975]; a three-dimensional spherical shell of occupations [Burton, 1972]; or a circumplex of facial expressions showing mixed emotional states [Nummenmaa, 1990]. This "centrifugal effect" is also produced when applying vote-counting to incomplete H-sorts [Shweder, 1972, shown in White, 1978]. Sometimes these two artifacts are present simultaneously and one sees a configuration where tight clusters of stimuli are spaced at intervals around a circle or on a sphere, to maximise their mutual separation.

In such a situation where the estimates of large d_{ij} are noisy, lowest Stress values are obtained when the elements are arranged in a hyperspherical shell in the available dimensionality, maximising the distances between them in an approximation of the degenerate solution in which they are mutually equidistant and Stress drops to zero. The more dimensions the better the approximation. Characteristic of this effect is the absence of a sharp 'elbow', indicating the optimal number of dimensions, in the plot of Stress against dimension. Russell [1983] comments on this absence in two out of seven cases where non-English words for emotions were sorted.

Rao and Katz observed spurious dimensions appearing in their MDS analysis of synthetic co-occurence matrices. Many researchers found sorting co-occurences too noisy for MDS, demanding an implausible number of dimensions to bring the Stress down to acceptable levels, and confined themselves to clustering analysis [Jones & Ashmore, 1973; Miller, 1969; Reading, Everitt & Sledmere, 1982].

My chief concern is that less cautious researchers will interpret a scaling solution, artifacts and all, and misinterpret artifacts as genuine features of the stimulus space (for instance, interpreting both dimensions of a horse-shoe created from a one-dimensional set of stimuli, such as the 22 pain descriptors scaled by Morley [1989]). Note also that the similarity between MDS solutions (as in White's [1978] cross-cultural comparison) is exaggerated by an artifact shared in common.

Variant vote-count distance estimates

Several researchers have suggested modifications to the way the co-occurence matrix is compiled, before scaling it. In all of these variations on vote-counting, the position of an element in stimulus space still affects the estimated distances between two elements nearby, which is undesirable, as Torgerson argued in his critique of vote-counting in the context of triads.

Ward [1977] proposed weighting the m-th sort's contribution to E by G(m), the number of groups the elements were sorted into. The fact that i and j have been sorted together is more informative when there are more groups, because the opportunities to place them in different piles are greater (this is still treating group co-occurence as a form of low-resolution proximity data. Preferentially weighting the sorts with more and smaller groups is a way of placing higher emphasis on proximities obtained with smaller þ_m). Russell and his colleagues have used this weighting scheme extensively. They also followed Ward by using a procedure in which subjects are asked to repeatedly sort the stimuli, with the numbers of groups set in advance.

This is only a first approximation, which breaks down at extremes. To split N stimuli into N single-stimulus groups conveys no more information than lumping them into one pile. Burton [1975] proposed a more sophisticated weighting scheme which modulates a sort's contribution to the dissimilarities of dyads (i,j) (i and j grouped together) and (i,k) (i and k in different groups):

If subject m partitions the element set into $\wp(m) = \{ I_1, I_2, ...I_{G(m)} \}$, the probability of finding any two elements chosen at random in pile I is

$$H_I = \binom{N_I}{2} \Big/ \binom{N}{2}$$

and $Q_m = 1 - \sum_{I \in \wp(m)} H_I$

is the probability that the random elements will be in different piles. Burton defines a distance estimate

$$z_{ij} = \sum_m \delta_{m,ij}$$

$$\text{where } \delta_{m,ij} = \begin{cases} 0 & \text{if } i = j \\ C + \log_2 H_I & \text{if } i,j \in I \\ C - \log_2 Q_m & \text{if } i,j \text{ in different piles} \end{cases}$$

where $C = 2 \log_2 \{N! / (2 (N\text{-}2)!) + 1\}$

Thus $\mathbf{z}_{ij}$ takes into account both the number of groups and their sizes. Co-occurence of i and j in a small group is less likely, and indicative of greater similarity, than if the group were larger. This is reasonable, given the correlation I noted previously between $|I|$ and average d_{ij} (for $i,j \in I$).

Using occupations and behaviours as stimuli, Burton found that his weighted matrix Z produced more plausible reconstructions than the standard co-occurence E, though at the expense of higher Stress values. Burton also tried a third measure G to which larger groups contributed more than smaller groups; it produced the worst results of the three (in terms of forming tight clusters).

Finally, there is a measure of dissimilarity which Rosenberg calls the *delta* transform (besides Rosenberg and his co-workers, see Jones and Ashmore [1973]; Jones, Sensenig and Ashmore [1978]; Paddick [1978]). The transformed estimate of distance $\delta^{(2)}{}_{ij}$ is a function of the averaged co-occurence matrix which takes into account *indirect* links between i and j (i.e. elements k which bridge i and j by frequently co-occuring with one or the other), as well as direct links:

$$\delta^{(2)}{}_{ij} = \Big(\sum_{k=1}^{N}(e_{ik} - e_{jk})^2\Big)^{1/2}$$

In effect this is treating rows of E as profile data (scale ratings), and computing Euclidean distances between them.

Drasgow and Jones [1979] proceeded from Monte Carlo experiments to the conclusion that the delta transform reduces Stress without improving the interpretability of dimensions. They also found the untransformed matrix to be more closely related to the underlying configuration than $\Delta^{(2)}$. This seems to be because the transform is non-monotonic, increasing large dissimilarities while small ones become smaller. The clusters which have a tendency to appear in co-occurence data, because of indirect links between elements, thus become tighter (reducing Stress in the process – recall that degenerate configurations are Stress-free). Van der Kloot and van Herk observed this unwanted clustering when they scaled both E and $\Delta^{(2)}(E)$ in their study [1991] of personality-trait adjectives and verbs of "getting one's way", data we encountered above.

Incidentally, the Block-model method of cluster analysis [Breiger, Boorman & Arabie, 1975; also McQuitty & Clark, 1968] uses a similar transformation, with the express purpose of bunching elements together; a proximity matrix is replaced with the *correlations* between its columns, transforming the matrix iteratively until it is reduced to 1s and 0s.

Individual differences

One digression finishes and another begins immediately: this time on the subject of individual variations. Examining a representative specimen of sorting data, one is struck by the range of subjects' sorts. Can all the differences be attributed to random fluctuations of a mental map (shared by all subjects) of the stimuli – the kind of fluctuations required if sorting data are to be scaled at all? Or can they be traced to systematic differences between subjects? Differences could take the form of continua – the INDSCAL and Points-of-View models; another possibility is distinct sub-populations of subjects, each sub-population having their own (shared) mental map.

Certainly the process I have described (reconstructing the map on the assumption that each sort is a slice through a clustering dendrogram) raises no barriers to fitting individual dimensional weights. But I do not expect such weights to be meaningful. There is simply not enough information in each sort, too few comparisons involving large dissimilarities.

Applying the INDSCAL option of MTRIAD to the van der Kloot-van Herk data, the algorithm takes advantage of the additional degrees of freedom to increase the solution's Likelihood. But crucially, the axes of figure 7.17 are arbitrary. One can rotate the solution through 45° and the program achieves a comparable increase in Likelihood by assigning a different set of dimensional saliences. Unlike the situation when processing H-sorts, there is no single optimal alignment for the program to rotate the solution to.

In this example, independent values are available for the subjects' dimensional weights: each subject also assessed the dissimilarities between pairs of stimuli directly. The sorting-data weights and those produced by scaling these dissimilarities show no significant correlation.

The same comparison is possible with the POOC material, where there are 33 respondents who provided both sorting and pairwise data. The stimulus sets are of different sizes, with little overlap between them, but they inhabit the same 'occupation space', and dimensional weights should be comparable between them. However, there is no correlation between the two sets of dimensional weights for each correspondent. The conclusion is inescapable that if systematic variations exist among people's strategies for sorting stimuli, then the nonsystematic variation (noise) drowns them out.

Little success has been reported from applying the INDSCAL model directly to the C_m matrices. Richards and Hanson [1985] found no differences in how American Sign Language letter-signs were sorted by native and second-language signers. Turk, Wack and Kerns [1985] compared how two sets of health professionals sorted "pain behaviours", and found no differences.

This does not rule out the possibility of detecting individual differences in other ways, independently of reconstructing a map.

There have been a number of attempts to derive meaningful indices of personality variation from sorting data, without scaling them. They are outside the scope of this thesis, but in passing I note a study [Arabie & Boorman, 1973] which examined the question of distinguishing different "cognitive styles" in the ways subjects sorted stimuli. Coxon and Jones [1979a] looked at the 'height' of a sort, as a reflection of the level of semantic generality for that subject's representation.

More to the point is a paper by Hubert and Levin [1976]. This describes statistics for distinguishing whether two sorts differ significantly. Is the difference between a subject's sorting performance, and an expected partition (or more generally, a proximity matrix), greater than chance can account for? (Bersted, Brown and Evans [1970] considered the same problem).

This suggests the idea of defining a function to quantify the dissimilarity between pairs of partitions. Given a table of 'distances' between sorts, MDS on the space of *subjects* becomes possible (a Q-analysis, as opposed to the P-analysis performed on *stimulus* space).

Numerous distance functions have been proposed. Arabie and Boorman [1973] tested 12 of them, using simulated data (see also Boorman and Arabie [1972]). The simplest function for sort distance – "pairbonds" – seemed better than more complicated information-theoretic measures. A later survey is Hubert and Arabie [1985]. Many of the functions they consider are generalised correlations while others are generalisations of the "profile distance" (differing in the details of correction or normalisation). Either way, the co-occurences are treated as distance matrices.

A problem common to all such functions arises when sorts with different G(*m*) are compared. How dissimilar are the partitions (1,2)(3,4,5) and (1,2)(3)(4,5)? The unequal numbers of groups leaves open the possibility that they come from the same mental map; the same dendrogram, sliced at different heights; in which case they are not dissimilar at all. But (1,2)(3,4,5) is equally compatible with a third partition, (1,2)(3,4)(5), which clearly *is* dissimilar to (1,2)(3)(4,5). Perhaps one should set the value of G(*m*) in advance, and only compare dendrograms sliced at the same height.

As it is, a distance function's dependence on variations in G(*m*) (unavoidable if it is to be a metric) dominates the scaling solutions, forming a primary dimension, to which other forms of variation are secondary. See, for example, the scaling of the Struhsaker data for vervet monkey sleeping groups, in Arabie and Boorman [1973]. The sorts with high G(*m*) form a tight central cluster, round which other sorts are arranged in concentric circles of diminishing G(*m*).

To summarise studies in which sorting data were Q-scaled, the results were negative, whenever external criteria were available with which to assess the meaningfulness of the scaling solution. In a study in which psychologists sorted the names of influential figures in psychology, Rosenberg and Gara [1983] performed a Q-scaling on their respondents, anticipating a connection between sorting behaviour and professional affiliation, but found none.

I suspect that if information about how individual subjects differ is to be derived, the configuration of stimuli must be scaled in parallel. It is easier to compare how two sorts differ from a configuration (their degrees of incompatibility) than to compare the compatibility *between* them (one can readily construct situations where seemingly quite different partitions result from two subjects viewing a configuration with only slight variations in their dimensional saliences).

It may be that the residual-forces form of Q-analysis presented in Chapter 4 is still useful. If nothing else, it is potentially a way to identify sub-populations (latent classes) among the sorters, if the differences between sub-population configurations are larger than in the cases we have encountered.

Conclusions

The sorting procedure has proved its value many times over since it was introduced. In situations involving large numbers of stimuli, no other procedure is practical. The hierarchical sorting method provides richer information – the example using D15 stimuli shows that it allows the discrimination of individual variations – and deserves to be used more widely. Large element sets can be handled by the H-sort method by having each subject H-sort a subset of the total elements, with the subsets overlapping. In the examples I have described, each subject sorted two subsets, between them containing all the elements, but this is not essential. In an analogous case, using free sorting, Kraus, Schild and Hodge [1978] arranged their 220 items (occupation titles) into four overlapping subsets of 90 items, with subjects sorting one subset each.

However, the analysis of partitioned sorts and H-sorts proves to be beyond the capacity of the vote-counting method – too much "noise" is added to the aggregated co-occurence matrices. On top of this there are the characteristic artifacts which vote-counting introduces. I argue that there is room for improvement in the analysis of sorting data, and propose a "reconstructed dyad" method.

The analyst must bear in mind the limitations of the sorting and hierarchical methods (notably, the paucity of the large-dissimilarity comparisons which convey the global structure of the perceptual space) and eschew forms of analysis which purport to extract more information than warranted by the observations.

For mapping large sets of stimuli there is a lot to be said for augmenting sorting data with triads or pairwise comparisons for a limited subset of them, providing an "armature" and relying on local information to fill in the gaps between them with the remaining stimuli. A hypothetical example might involve combining the sorting data for 24 faces from the Lightfoot series with Gladstones' rigorous treatment, limited to 10 faces by the restrictions of the triadic method. "Scaling methods [triadic or pairwise data] should probably be reserved for those cases where we want a particularly accurate study of a relatively small number of items" [Miller, 1970, p. 571].

Another possibility is combining sorts with scale ratings to provide the global structure. There are many cases where ratings were elicited and preference-mapped *a posteriori* to the configuration, to help interpret the axes [e.g. Burton, 1972; Jones, Sensenig & Ashmore, 1978].[3] Since the configuration is distorted by precisely the lack of the information which the ratings contain, this seems less than optimal. I argue for joint scaling of the scale and sorting data, as in the woodblock example. Each form of data makes up for the other's deficiencies.

The same argument applies to other forms of data deficient in large-dissimilarity comparisons, such as the pick *k/N* form.

---

[3] Schmidt [1972] presents an interesting variant: instead of individual stimuli, subjects were asked to rank the clusters they had formed, on a scale of "degree of agreement with". Similarly, Kraus *et al* [1978] asked subjects to rank their clusters of occupation titles according to "social standing".

8. TRIADS REVISITED – ASPECTS OF DATA COLLECTION

A theme of this thesis has been the problem of constructing large-N well-landmarked perceptual maps when the stimuli are not visual, making it impossible to present them simultaneously in order to sort them. This brings me back to the Method of Triads.

The size of the list of all possible triads is the factor limiting applications of the Triadic methods. Dong [1983] found that subjects' boredom and fatigue were already increasing the fallibility of their responses, for as few as nine stimuli (84 triads), though not systematically enough to distort the MDS solution. With more elements, the number of triads, proliferating as N^3, soon grows prohibitively large.

Semantic and non-visual stimuli, with their requirements of conscious processing and internal representation, seem to be more tiring than visual ones. The gestalt-forming, parallel-processing features of vision made 2240 triadic comparisons between colours acceptable [Stalmeier & de Weert, 1991a, 1991b], and even 83 720 triads, in a truly heroic study of visual textures [Harvey & Gervais, 1981].

Experiments with more than 11 or 12 non-visual stimuli can only proceed by winnowing down the list of triads to be presented. Which triads to include? I will review previous answers to this question, before describing a new approach.

Balanced Incomplete Designs

A BID is a list of $N\,(N\text{-}1)\,\lambda$ triads in which each dyad (i,j) appears λ times, where the constant λ is at least 1 and at most N-2. To scale $N = 17$ odours, MacRae, Rawcliffe, Howgate and Geelhoed [1992] used a $\lambda = 3$ design, amounting to 20% of the complete list of triads. Balanced Incomplete Designs have also been used by Arabie, Kosslyn and Nelson [1975] (where $N = 12$ and $\lambda = 3$), the POOC [Coxon *et al*, 1975], and Kirk and Burton [1977] ($N = 13$, $\lambda = 4$). Burton and Nerlove [1976] describe the combinatorial principles involved in the construction of BIDs, and provide examples for various $N < 21$ and for a variety of λ values.

There is an analogy with balanced incomplete designs for pairwise dissimilarity data: designs which omit dissimilarities so that for each i, δ_{ij} is known for μ values of j, μ being fixed. Often these are cyclic. See Spence and Damoney [1974]; Giraud and Cliff [1976].

I will describe the construction of a BID for a case not covered by Burton and Nerlove, that of $N = 12$ and $\lambda = 4$. In what follows, the stimuli will be labelled with integers from 1 to N.

Begin with the triad <1,3,7> plus its eleven cyclically incremented equivalents, $<1 \oplus i, 3 \oplus i, 7 \oplus i>$ $(0 < i < 12)$, where $\oplus$ is addition (modulo N). Let the cyclic increments be implicit and represent the whole family as {<1,3,7>}. The object is a set of triad families in which each of the 66 dyads of the complete connected graph appears four times.

It is easy to check that the seven triad families, {<1,3,7>}, {<1,2,7>}, {<1,2,4>}, {<1,5,10>}, {<1,6,10>}, {<1,3,6>}, {<1,2,3>}, plus one of the form $\{<1 \oplus i, 5 \oplus i, 9 \oplus i>\}$ where $0 \le i < 4$ add up to 4 copies of the complete connected graph, so the eight triad families form a BID.

Tinkering with lines on clockfaces for $N = 15$, I found three solutions for $\lambda = 3$ (105 triads each). They contain the triad families:

- solution 1: {<1,3,7>}, {<1,4,8>}, {<1,2,3>}, {<1,5,10>}, {<1,4,9>}, {<1,3,6>},{<1,2,8>}
- solution 2: {<1,5,7>}, {<1,5,9>}, {<1,6,7>}, {<1,6,9>}, {<1,4,6>}, {<1,7,8>}, {<1,2,4>}
- solution 3: {<1,5,6>}, {<1,3,4>}, {<1,5,8>}, {<1,3,9>}, {<1,3,8>}, {<1,2,6>},{<1,7,10>}

Any two of these combine to form a $\lambda = 6$ design, while the sum of all three is a $\lambda = 9$ design, which is mainly of interest because it provides a solution for $\lambda = 4$, which is simply the set of all triads *not* included in the $\lambda = 9$ design.

Finally, an unpublished solution for $N = 12$, $\lambda = 1$ comprises 22 triads, 6 in each of the families $\{<1 \oplus_{12} 2i, 2 \oplus_{12} 2i, 3 \oplus_{12} 2i>\}$, $\{<2 \oplus_{12} 2i, 4 \oplus_{12} 2i, 9 \oplus_{12} 2i>\}$ and

$\{<1 \oplus_{12} 2i, 4 \oplus_{12} 2i, 7 \oplus_{12} 2i>\}$ for $0 \le i \le 5$,

plus for of the form $\{<1 \oplus_{12} i, 5 \oplus_{12} i, 9 \oplus_{12} i>\}$ for $0 \le i \le 3$.

The case of $\lambda = 9$, $N = 15$ is an example of a Balanced Incomplete Block Design (BIBD). The special feature of a BIBD is that it consists of several blocks of triads, each being a BID in its own right: the separate $\lambda = 3$ designs, in this case. No two blocks contain triads in common.

BIBDs are a convenient way of sharing out a balanced set of triads among multiple subjects, to reduce the individual workloads, or to cover a wider range of triads without including any one triad repeatedly. Thus, a study of 15 musical intervals [Levelt, van der Geer & Plomp, 1966] and another on the effects of phase differences on timbre [Plomp & Steeneken, 1969] both used a BIBD with $\lambda = 4$, composed of 4 $\lambda = 1$ BIDs, so that each subject made judgments on 35 triads. This is different from the $\lambda = 4$ BID described earlier.[1]

Burton and Nerlove concentrate on BIBDs. For $N = 13$, they provide two BIDs for $\lambda = 1$, which together constitute a $\lambda = 2$ BIBD:

1. {<1,4,5>} and {<1,6,8>}
2. {<1,2,5>} and {<1,3,8>}

I found two more BIDs, which combine with the first two, forming the BIBD for $\lambda = 4$:

3. {<1,2,4>} and {<1,3,6>}
4. {<1,5,10>} and {<1,2,8>}

Similarly, Burton and Nerlove provide a $\lambda = 2$ BIBD for N = 19, made up of two $\lambda = 1$ BIDs:

1. {<1,3,6>}, {<1,8,9>}, {<1,5,11>}
2. {<1,2,6>}, {<1,4,11>}, {<1,3,9>}

I augmented these with another two, forming the $\lambda = 4$ BIBD:

3. {<1,4,6>}, {<1,2,9>}, {<1,7,11>}
4. {<1,5,6>}, {<1,8,11>}, {<1,7,9>}

---

[1] In the domain of pairwise dissimilarities it is known [Spence & Domoney, 1974] that some partial designs are more "effective" than others. Given two incomplete sets of δ_{ij}, both balanced with the same μ, the better design for recovering the configuration design is generally the one containing fewest triplets δ_{ij}, δ_{jk}, δ_{ik}. It is a moot point whether similar statements can be made about rival BIDs having the same value of λ.

However, there is no compelling reason to balance a data set at the level of individuals (other than habits of thought inherited from factorial experiment designs in agriculture, along with the terminology of "blocks"). When MacRae *et al* found 8 assessors willing to judge the 136 triads constituting a $\lambda = 3$ BID for 17 odour stimuli, they simply shared the triads out randomly, 17 per assessor, unconcerned as to whether each assessor judged each dyad the same number of times.

Unbalanced Designs

It is interesting to note that the designs of a number of experiments which the researchers believed to be balanced were in fact unbalanced [Burton & Nerlove; Kirk & Burton; etc] because they used the odd-one-out variant of the triadic method. When one picks i as the odd one out of $\langle i,j,k \rangle$, this provides the information that $(i,j) \gg (j,k)$, and $(i,k) \gg (j,k)$. Thus (j,k) takes part in two comparisons; (i,j) and (i,k) are participants in only one each, reducing the number of judgments – which is, after all, the point of the variant. First described by Andrews and Ray, this variant has been widely used in anthropological studies: there is no option in the triads-analysis section of the ANTHROPAC software [Borgatti, 1991] for any other type.

I draw two conclusions from this. First, shorter distances are specified more reliably by odd-one-out data: the further j lies from i, the fewer inequalities constraining the value of d_{ij}. In view of what is known about the important role of large dissimilarities, this argues for caution in acquiring and scaling odd-one-out triads. Chapter 3 made the same point. Ideally, stimuli will be scattered evenly enough for small-scale relatively precise maps of portions of the perceptual space to fit together and recover the space's global structure despite the paucity of long-distance comparisons. Elongated configurations are bad.

The second conclusion is that incomplete data do not have to be balanced. Though unbalanced data contain more comparisons for some d_{ij} than others, this may not be a bad thing. The varying number of comparisons should be borne in mind if the data are vote-counted:

$$vc_{ij} = \sum_{(k,l)} \varepsilon_{ij,kl} / \sum_{(k,l)} (\varepsilon_{ij,kl} + \varepsilon_{kl,ij}) \tag{2.12}$$

An example of an unbalanced incomplete design in Bechtel's study of nine Munsell chips. Recall that three chips were made "standards", the other six "comparison" stimuli. Instead of 84 triads, Bechtel's design used 45: each containing one standard and two comparison stimuli.

It is hard to see why balance of the data should make any difference when MDS is performed directly on the dyad inequalitites. In Monte Carlo simulations, for the more general case of tetradic comparisons, Bissett and Schneider [1992] found no difference between BIDs and the same number of comparisons selected at random. But when Takane performed similar simulations [1978] as part of evaluating his MLE analysis of tetradic comparisons, he found a complete (hence balanced) set of triads to give more accurate recoveries than an equally large but randomised set (presumably unbalanced) of tetrads.

However, we saw in Chapter 5 that Stress is not wholly adequate as a badness-of-fit function. A sufficient level of incompleteness in the data unmasks its potential for distortions and artifacts. I have gained the impression that data which are unbalanced as well as incomplete become more vulnerable to these artifacts. The remedy is to switch from minimising Stress to maximising Likelihood.

I argue now that some triads are more informative than others. If a tentative sketch of the configuration is available (perhaps the perceptual space has been coarsely mapped by scales, or triads from previous subjects are available), it becomes possible to omit triads for which the judgments can be predicted with reasonable confidence, and to concentrate on presenting others for which the expected information content is greater.

Consider the approach taken by Wright [1965] with an element set of 17 colours. Wright divided the region of colour space occupied by the colours into 6 compact, overlapping sub-regions, each containing 7 elements (many were shared between sub-regions), and judged only those triads where *i,j,k* belonged to the same sub-region. As well as reducing the number of triads to 210 (6 x 35), this restricts the range of distances being compared, excluding many elongated triads of low information content, where two elements belong to one subset while the third, in a different sub-region, is the obvious one out.

Another unbalanced design features in Krantz [1967]. Again, the stimuli were colours. The design consisted of comparisons between dissimilarities expected to be roughly equal; some triads comparing large dissimilarities, others comparing small ones. Krantz was using Torgerson's analysis, in which there is no point comparing highly unequal dissimilarities, since the conversion of p (the proportion of replications in which $(i,j) \gg (j,k)$) into an equation, $d_{ij} = d_{jk} + a(p)$, becomes impossible if p is 0% or 100%.

For tetradic data, Bissett and Schneider noted that many comparisons could be omitted since the responses were predictable from other comparisons. However, their argument only applies to scaling in one dimension.

My final example is the Interactive Similarity Ordering method (ISO) described by Young, Null & Sarle [1978]. Multiple-choice questions are presented to the subject, who must choose which of a list of stimuli is most similar to a target stimulus. When the length of the lists is restricted to two, this procedure reduces to triadic comparisons. Triads are omitted when the response can be predicted from previous responses (assuming them to be transitive).

Interactive Incomplete Designs

In this context, the pertinent feature of the ISO procedure is its interactivity. Triads are selected on the basis of the subject's responses to earlier triads.

With interactive selection it is possible to map a perceptual space using fewer triads than a BID. I will describe an interactive procedure, implemented in a computer program "Triskele", which selects $O(N)$ triads to map N stimuli (as opposed to the $O(N^2)$ required by BIDs, for constant λ).

Reduced to the broadest outlines, the Triskele procedure consists of two phases. In the first phase, a basis or skeletal set of N_b stimuli ($8 \leq N_b \leq 10$) is mapped, using a BID with λ_b or a complete set of triads (of the "triadic combinations" form). This phase is non-interactive but it provides a framework within which the remaining $N_r = N - N_b$ stimuli can be located interactively.

The second phase is repeated for each of the remaining N_r stimuli. The l-th stimulus ($N_b < l \leq N$) is located using triads $<i,j,l>$, for $i,j < l$. I divide this second phase into phases 2a and 2b. 2b consists of presenting the subject with triads chosen for high expected information content, having the maximum potential to refine the provisional position x_l. The information gained from each triad is incorporated in the map by iterating the optimising algorithm: downhill descent, in Triskele. Thus each triad affects the information content expected from subsequent ones, and the positions x_i and x_j are refined as a side-effect of locating l.

An initial x_l must somehow be obtained. That is the role of phase 2a, which presents a small number (typically 6 to 9) of triads $<i,j,l>$ where nothing is assumed about x_l except that it is somewhere in the area of perceptual space spanned by the previous l-1 elements. In this situation, good criteria for selecting $<i,j,l>$ are that x_i and x_j should be neither particularly central nor excessively peripheral, with above-average d_{ij}.

2a can be omitted if tentative positions for the remaining elements are available from some other source (e.g. scale ratings or another subject's perceptual map, or theoretical considerations).

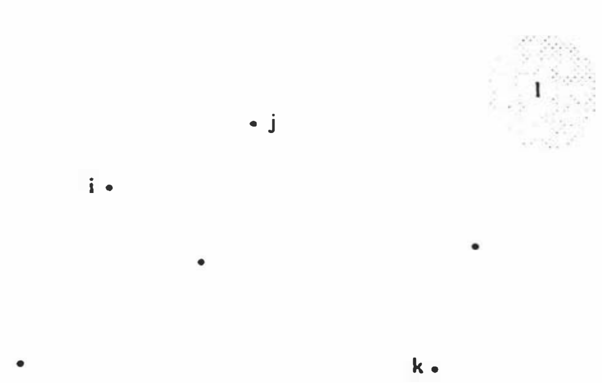
Note the contrast with BIDs. Far from being balanced, a Triskele session involves the various dyads in different numbers of triads, ranging down to zero. It is not necessary for the majority of (i,j) to take part in triads at all, and if they did, a N^2 term would dominate the size of the triad list.

The Triskele procedure differs from ISO in its use of a provisional configuration, constructed from triads so far, as a basis for selecting subsequent ones. In this, Triskele has much in common with the ISIS procedure (Interactive Scaling of Individual Subjects) [Giraud & Cliff, 1976; Young & Cliff, 1972], which has been implemented in two programs, ISIS and INTERSCAL [Cliff, Giraud, Green, Kehoe & Doherty, 1977; Kehoe & Reynolds, 1977]. Both programs begin by eliciting enough data to construct a configuration for a basis set of stimuli. The principle underlying ISIS is that in a P-dimensional space, a point can be located by specifying the distances from that point to (P+1) known points. For the l-th stimulus ($N_b < l \leq N$), it prompts the subject for at least (P+1) dissimilarities δ_{jl}, where the "benchmark" points x_j are chosen for their peripheral positions in the configuration, thereby maximising their effectiveness for triangulating x_l (if x_l is sufficiently peripheral, it can become a benchmark for locating subsequent stimuli). For ISIS to proceed past the first phase, a value must be assigned to P. Triskele shares this requirement.

These early forms of ISIS are metric. They expect ratio-level data. A non-metric version of the ISIS procedure has been described [Hamer, 1981], closer to Triskele. Hamer's ambitious extension was also capable of supplementing the judgments from the current subject with data elicited from previous judges, as an aid to selecting the benchmark items j, while allowing for individual variations between the current and previous subjects. Despite these features, Hamer found no significant improvement in the procedure's performance.

To explain the logic of Triskele in more detail, I should quantify the vaguely-worded descriptions of some triads being more "informative" than others.

Figure 8.1 Triad involving new element, l



Imagine a situation such as figure 8.1, where precise postions x_i, x_j, x_k have already been found (somehow) for three of the elements, while element l is in the process of being located, a process that has provisionally localised x_l to somewhere in a circular region, L (cross-hatched). L could be a 90% confidence region, or 95%; the exact figure, and the exact radius of the circle, are not crucial in this argument. There is no point in presenting the subject with <*i,j,l*>, since the response is unlikely to come as a surprise; it is likely to be

$$\delta_{ij} < \delta_{il}, \qquad \delta_{ij} < \delta_{jl}, \qquad \delta_{il} > \delta_{jl},$$

which imposes no fresh constraints on x_l. $<j, k, l>$ would be more informative (in the sense of telling the analyst something not known already).

In the comparison between δ_{ij} and δ_{il}, let < and > stand for the responses $\delta_{ij} < \delta_{il}$ and $\delta_{ij} > \delta_{il}$ respectively. Let p = Pr(<) and q = Pr(>) = 1 - p. The information value of < and > are -$\log_2(p)$ and -$\log_2(q)$ respectively. Thus, the expectation value of the information received from that comparison is

$$\mathrm{I}(i,j,i,l) = -p \log_2(p) - q \log_2(q) \tag{8.1}$$

As in Chapter 5, I approximate the probabilities with logistic functions:

$$\Pr(<) = \{1 + \exp(\tau\, D_{ij,il})\}^{-1}, \qquad \Pr(>) = \{1 + \exp(-\tau\, D_{ij,il})\}^{-1},$$

where $D_{ijil} = d_{ij} - d_{il}$, the difference in reconstructed distance, while τ subsumes all contributions to the uncertainty of the response: the small uncertainties in x_i and x_j (shown as zero in figure 8.3), the large uncertainty in x_l, plus subject error. The approximation is a crude one, concealing a number of assumptions, but it suffices.

$\mathrm{I}(i,j,i,l)$ becomes a roughly Gaussian function of $\Delta_{ij,il}$, with a maximum of 1 at $D_{ij,il} = 0$, and standard deviation inversely proportional to τ.

The same argument applies to the comparisons between (i,j) and (j,l), and (i,l) and (j,l). The uncertainties are different, since L is not necessarily circular, but one can still argue that the expected information content is maximised when the distances being compared are equal. Let $\mathrm{I}(i,j,l)$ be the total expected information from that triad, $\mathrm{I}(i,j,l) = \mathrm{I}(i,j,i,l) + \mathrm{I}(i,j,j,l) + \mathrm{I}(i,l,j,l)$. This treats the comparisons as independent – as if presented with the Torgerson method of "complete triads" – which introduces another approximation.

I conclude that the most informative triad is one forming an equilateral triangle in the provisional reconstruction of perceptual space.

An alternative to maximising expected information, is to select those triads for which the expected contribution to Likelihood is most negative. I liken the problem of interactively locating l to a process of forming a hypothesis about x_l, which one then attempts to falsify, in order to replace it with an improved hypothesis. Another rationale for selecting Likelihood-minimising triads is that they are the ones from which one expects the greatest contributions to forces on the elements, providing the strongest constraints on their positions

The expectation value of the contribution to log Likelihood from comparing (i,j) against (i,l) follows from the definition of log Likelihood:

$$I^*(i,j,i,l) = p \ln(\{1 + \exp(\tau D_{ij,il})\}^{-1}) + q \ln(\{1 + \exp(-\tau D_{ij,il})\}^{-1}) \qquad (8.2)$$

$$= I(i,j,i,l) \text{ times a constant.}$$

These are all approximations to a rigorous approach to triad selection, which would set confidence ellipsoids for the stimuli [see Ramsay, 1978], and for each $\langle i,j,l \rangle$, calculate the expectation values for reductions in their extents, finding the (i,j) which effect the greatest shrinkage of the ellipsoids. Equivalently, one might envisage this as a process of shrinking the $(N \cdot P)$-dimensional confidence ellipsoid around the single point X in configuration space, an ellipsoid which is most elongated along axes corresponding to the positions of elements which have not yet been located.

In practice Triskele uses a simpler criterion. It chooses the $\langle i,j,l \rangle$ which is closest to being equilateral, i.e. which has minimum

$$|d_{ij} - d_{il}| + |d_{il} - d_{jl}| + |d_{ij} - d_{jl}| \qquad = \qquad |D_{ij,il}| + |D_{il,jl}| + |D_{ij,jl}|$$

(as long as that triad has not been previously used). Though not optimal, this criterion is fast. Sometimes it results in the selection of triads which have the effect of locating x_l in the direction of the short axis of its confidence ellipsoid, while ignoring the looser constraints on x_l in the direction of the long axis.

Selecting triads to minimise Likelihood has a drawback. Low Likelihood in a finalised configuration (or high Stress) could be caused by noisy data, too few dimensions, or simply a felicitous choice of triads. Dependent as they are on the particular triad choice, Stresses and

Likelihoods cannot be compared between Triskele sessions; nor do the usual tests of dimensionality apply. It remains to be seen whether τ remain sufficiently independent of triad choice for it to be comparable across sessions.

Returning to phase 2a, the triads presented therein are complete stabs in the dark. To change our metaphors for a moment, if we think of the configuration explored so far as an atlas of stimuli, the rôle of these preliminary triads is to determine the page on which x_l lies, before the 2b triads localise it more precisely. Triskele presents triads in groups of three, $<i,j,l>$, $<i,k,l>$, $<j,k,l>$, after considering all triplets of items i,j,k in search of the three which mark out a triangle with the greatest area. Barring pathological configurations, this ensures that i,j and k are peripheral, but not outliers, and that their centre of gravity is more or less central, and that d_{ij}, d_{jk}, d_{ik} are above average. Triads selected by these guidelines seem reasonably effective at triangulating x_l, whereever in the configuration it happens to be. This is repeated for further triplets until the specified number of phase 2a triads has been made up. Iterations of the downhill-descent algorithm follow, to integrate the responses into the configuration.

This is the process for $P = 2$. If, instead, the dimensionality is assumed to be 3, the program chooses quartets of elements, h,i,j,k, forming a tetrahedron of maximum volume, and presents $<h,i,l>$, $<i,j,l>$, $<j,k,l>$ and $<k,h,l>$. The generalisation to higher dimensionalities is straightforward though not yet implemented in Triskele.

Monte Carlo simulations comparing the Triskele approach against BIDs (analysed with vote-counting) indicate that both deteriorate with the decline in the amount of information as λ is decreased. However, BIDs deteriorate faster; they are vulnerable to the fluctuations caused by comparisons between (i,j) and atypically close and atypically distant pairs.

I also conducted Monte Carlo experiments using the Stalmeier-de Weert sets of triads to simulate a subject's responses instead of random numbers – consulting their data to see what a given subject's actual responses were to a given triad. The crucial test was whether Triskele's performance is superior to a Balanced Design involving the same number of triads.

I used two BIDs as the standards for comparison, with $\lambda = 2$ and 4, i.e. 80 and 160 triads. Three Triskele settings involving similar numbers of triads were as follows:

T1. 60 triads in phase 1 ($N_b = 9$, $\lambda_b = 5$); 14 triads to locate each of the remaining 7 elements (5 in phase 2a, 9 in phase 2b), for a total of 158 triads.

T2. 81 triads in phase 1 ($N_b = 11$, $\lambda_b = 5$); 14 triads to locate each of the remaining 5 elements (5 in phase 2a, 9 in phase 2b), for a total of 161 triads.

T3. 44 triads in phase 1 ($N_b = 12$, $\lambda_b = 2$); 9 triads to locate each of the remaining 4 elements (4 in phase 2a, 5 in phase 2b), for a total of 80 triads.

Entries in the table are the congruences between a reconstructed configuration based on partial data, and the complete-data configuration (see note 3, page 72).

Design:	$\lambda = 4$	T1	T2	$\lambda = 2$	T3
subject:ER	0.881	0.950	0.942	0.781	0.789
EU	0.958	0.962	0.976	0.900	0.884
GE	0.942	0.896	0.899	0.875	0.860
JA	0.983	0.989	0.991	0.828	0.975
MA	0.949	0.946	0.889	0.789	0.847
NA	0.964	0.987	0.979	0.892	0.948
PE	0.975	0.988	0.987	0.963	0.954
PI	0.970	0.970	0.980	0.901	0.864
RE	0.884	0.906	0.974	0.835	0.856
YV	0.954	0.979	0.967	0.947	0.920
average:	0.946	0.957	0.958	0.873	0.891

For some subjects, the Triskele strategies actually perform poorer than a BID. Nevertheless, on average the Triskele reconstructions are better. The differences may seem small, but it must be remembered that small changes in a congruence close to 1 corresponds to a large increment in accuracy.

A slightly different strategy is called for if the user opts to reduce the workload for subjects by asking for odd-one-out judgments only, instead of following the full Method of Triadic Combinations. Additional uncertainty is thereby introduced into the information content of the response. The response contains only two comparisons rather than three, with the subject deciding which two to include.

Instead of calculating $I(i,j,l)$ for each possible triad by weighting each of the three possible responses by the probability of its occurence, I programmed Triskele to expect the worst: following a minimax strategy, it assumes that the subject's response to $<i,j,l>$ will be the least informative (least surprising) one; in other words, the odd-one-out will always be the one predicted by the configuration. Thus Triskele selects (i,j) to minimise the maximum of $(|D_{ij,jl}| + |D_{il,jl}|, |D_{ij,il}| + |D_{il,jl}|, |D_{ij,il}| + |D_{ij,jl}|)$. So behind the selection of the next triad for presentation is the assumption that the subject will grudgingly respond in whatever way returns least information.[2]

Implementations of the Triskele procedure exist for PC and Mac computers.

PC-Triskele

Like MTRIAD, the PC incarnation is written in Turbo Pascal. The user interface routines (i.e. the sections handling interaction with the subject) are rudimentary. They are minor parts of the program, added almost as an after-thought, invoked by other parts which do all the real work of integrating previous triads into the configuration and selecting the next one.

The interface leans heavily on a geometrical metaphor. Stimuli are presented – in written form, if they are conceptual or semantic; played through the soundboard-amplifier combination, in the event of sounds; in either case, labeled with a 1, 2 or 3 – and a choice or choices made by typing the corresponding digit key. As a prompt for confirmation of judgments, the subject is presented with a scalene triangle, with the appropriate numbers at the corners; or an isosceles triangle if only primary judgments are being made.

The program reads N, N_b, the BID for N_b (if any), and the stimuli themselves (or identifying labels, if they are auditory) from an initialisation file. The first N_b stimuli in the list provide

---

[2] This pessimistic policy seems in keeping with the motto of the original Triskele: ***Quocunque Jeceris Stabit.***

the basis set; the order of the remaining ones determines the sequence in which they are incorporated into the configuration. An optional field in the file allows the experimenter to specify a BID to use with the basis set. When $N_b = N$, the basis set contains all the stimuli, and the interactive phases do not arise: this was the arrngement in three out of the four experiments in Chapter 3. In the experiment with $N_b < N$, I varied the order of stimuli for each subject, so that order effects, if any (fatigue, or the relative unfamiliarity of new items in comparisons with old ones) would not be systematic.

There is an option for specifying files, from which Triskele loads the triads from previous subjects' sessions, to provide a scaffolding for phase 2b triad selection and do away with phase 2a triads.

Tie-handling in Triadic Data

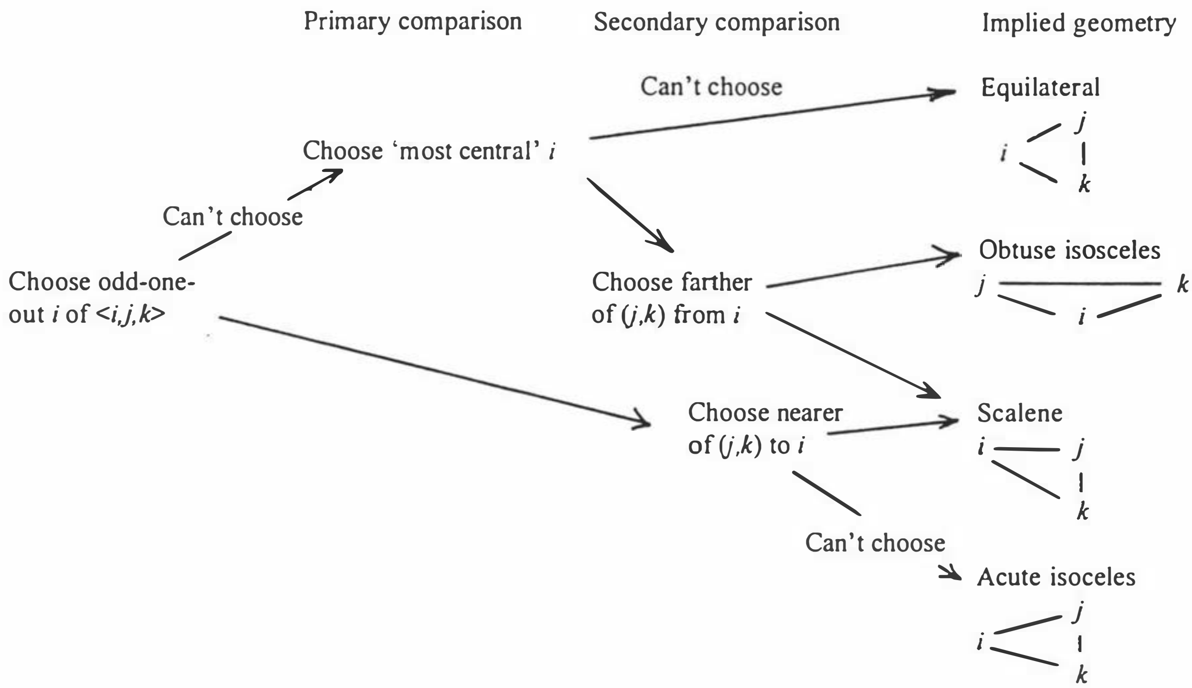
Tied comparisons are not normally encountered in triadic data; the usual experimental design does not offer subjects the option of saying that they cannot decide which of (i,k) and (j,k) is more dissimilar, i.e. which of i and j is more similar to k. It is a forced choice procedure.

Several participants in Triskele sessions volunteered the suggestion afterwards that for some triads, the difference between dissimilarities was small enough that effectively the choice between dyads was a random one, and that they would have preferred a "can't distinguish" option for such cases.

This option was easily incorporated in later versions of the PC-Triskele version. Subjects opt for "can't distinguish", at the primary or secondary judgment stage, by typing 0 instead or 1, 2 or 3. These choices have a contingent nature which most easily described as a tree (figure 8.2). The primary-comparisons-only variant omits the branch points pertaining to secondary comparisons.

At the terminus of each pathway through the tree, a geometry is implied for the arrangement of $<i,j,k>$ in perceptual space: a scalene or isosceles or equilateral triangle. An appropriate

Figure 8.2 Branching series of possible responses when ties are permitted



triangle is presented to the subject, along with a invitation to confirm the arrangement as an adequate summary of his or her perception of the dissimilarities, or to reject it.

Having provided this facility, I was faced with the question of how such choices should be included in analysis. It is convenient to treat the negative response, "can't distinguish", as equivalent to the positive statement that the two dissimilarities are equal (within a tolerance limited by the subject's discrimination and willingness to concentrate). Write this as $\delta_{ij} = \delta_{ik}$, or $(i,j) \approx (i,k)$.

In Coombs' taxonomy of data, tied comparisons are quadrant QIVb, as opposed to QIVa, which covers greater than/less than rankings. Data including both types of response are QIV.[3]

Kruskal proposed 'primary' and 'secondary' treatments of tied distance ratings – alternative terms are "weak" and "strong" ties. The primary treatment is to pretend that the distances

[3] Exclusively QIVb datasets are not unknown. In colour research they are sonetimes collected using a tetradic method [Indow & Aoki, 1983]: subjects indicate the dissimiliarity of two colours by choosing an equally dissimilar pair – these latter two both coming from a scale of greys.

were never compared, and to set $\varepsilon_{ij,ik} = \varepsilon_{ik,ij} = 0$. The tie is interpreted in a negative sense, as a lack of information about how d_{ij} and d_{ik} should be rank-ordered; the difference between them does not contribute to the configuration's badness-of-fit.

I am arguing that the secondary treatment is more appropriate. Here, any deviation away from strict equality between d_{ij} and d_{ik} contributes to Stress (and Likelihood). Each tie is expanded into a pair of inequalities: $(i,j) \gg (i,k)$; $(i,k) \gg (i,j)$. $\varepsilon_{ij,ik} = \varepsilon_{ik,ij} = 1$. This creates terms in $\partial L / \partial X$ corresponding to forces acting to minimise any such deviations as the hill-descent progresses.

In passing, I note that these seemingly dichotomous treatments are ends of a spectrum. More generally, for tied dissimilarities, the constraints of equality between the distances can be enforced with strictness sigma: $\varepsilon_{ij,ik} = \varepsilon_{ik,ij} = \sigma$ $(0 \leq \sigma \leq 1)$.

It seems preferable (or more productive) to allow a "can't distinguish" response, and interpret it as a strong tie, meaning that the difference between distances is small, than to demand that the subject tip the balance and make the less informative response that one distance is greater (with no indication of how much greater).

The effect of providing an option for ties is to expand the binary nature of the distance-comparison data into a ternary scale. My earlier reference to the subject's distance-discriminating tolerance glosses over the question of whether tied judgments correspond to a *range* of distance difference. A subject's responses might be governed by thresholds þ1 and þ2:

$$\begin{array}{ll} (i,j) \ll (i,k) & \text{if } \Delta_{ijik} \leq \text{þ}1 \\ (i,j) \approx (i,k) & \text{if þ}1 < \Delta_{ijik} < \text{þ}2 \\ (i,j) \gg (i,k) & \text{if þ}2 \leq \Delta_{ijik} \end{array} \tag{8.3}$$

where $\Delta_{ijik} = \delta_{ij} - \delta_{ik}$, and þ1 and þ2 are to be fitted to the data. Kruskal's secondary tie treatment is a special case of this with þ1 = þ2 = 0. If in fact þ2 - þ1 is large (i.e. the subject is using the tie option in excess), the assumption of strong ties introduces a possibility of distortion. Another possible subject response, based on the *ratios* of the dissimilarities, is

$$\begin{array}{ll} (i,j) \ll (i,k) & \text{if } \Delta_{ijik} / (\delta_{ij} + \delta_{ik}) \leq \text{þ}1 \\ (i,j) \approx (i,k) & \text{if þ}1 < \Delta_{ijik} / (\delta_{ij} + \delta_{ik}) < \text{þ}2 \\ (i,j) \gg (i,k) & \text{if þ}2 \leq \Delta_{ijik} / (\delta_{ij} + \delta_{ik}) \end{array} \tag{8.4}$$

I performed Monte Carlo simulations to assess this danger, simulating $\delta_{ij} = d_{ij} + e(\sigma)$. The limited discrimination of the hypothetical subject was modelled as in (8.4), with þ1 = -þ2 = þ. The criterion for recording a tie (to be expanded into inequalities) from the comparison between (i,j) and (k,l) was $|\Delta_{ijkl}| / (\delta_{ij} + \delta_{kl}) <$ þ. The Method of Triadic Comparisons was simulated, where it is not obvious how to simulate tied primary comparisons, since two distance comparisons are implicit, so the ties were only applied to secondary comparisons.

Each entry is the congruence coefficient between the true and the reconstructed configuration, averaged over 100 random three-dimensional configurations (with the three dimensions of equal salience). σ varied between 0 and C.4 (as a proportion of the average d_{ij} in a configuration). N was 15. The simulations involved a range of BIDs, with $2 \leq \lambda \leq 13$.

For þ = 0,

	$\sigma =$	0	0.1	0.2	0.3	0.4
λ =	2	0.967	0.964	0.943	0.915	0.881
	4	0.991	0.984	0.969	0.955	0.939
	6	0.995	0.991	0.982	0.971	0.961
	9	0.998	0.993	0.988	0.982	0.974
	13	0.999	0.996	0.988	0.982	0.974

For þ = 0.05 (affecting about 1/8 of secondary comparisons),

	$\sigma =$	0	0.1	0.2	0.3	0.4
λ =	2	0.976	0.964	0.947	0.919	0.881
	4	0.989	0.982	0.967	0.952	0.935
	6	0.991	0.986	0.979	0.966	0.955
	9	0.993	0.989	0.982	0.974	0.957
	13	0.995	9.991	0.982	0.974	0.971

For þ = 0.1 (affecting about 1/4 of secondary comparisons),

	$\sigma =$	0	0.1	0.2	0.3	0.4
λ =	2	0.974	0.967	0.947	0.918	0.887
	4	0.993	0.984	0.969	0.953	0.936
	6	0.995	0.990	0.981	0.971	0.957
	9	0.997	0.993	0.985	0.976	0.962
	13	0.998	0.994	0.984	0.978	0.972

Here we see the expected decline in accuracy of reconstruction as λ decreases and σ increases. For a high level of tied responses, we see a small degradation in the accuracy, especially when σ is high. However, for a more reasonable number of ties, the degradation is negligible; indeed, the performance is increased at low λ, which makes sense since more information – that the distances are close – is conveyed. These simulations were run using the Stress-minimising form of MDS, and it may be that MLE would make a difference.

Similar results arise, simulating the Complete Method of Triads. Not many more tied responses are introduced: the secondary comparisons, in the Method of Triadic Comparisons, include most of the problematical choices between dissimilarities of similar magnitude.

Other researchers have made the similar point that a tie between (i,j) and (i,k) should be interpreted and scaled as equivalent to the assertion that $|D_{ijik}|$ is less than the smallest difference between a pair of dyads which were *not* tied.

Another special form of triadic response which might be modeled as an expanded scale of dissimilarity difference is the confidence rating, encountered in chapters 3 and 5. Arguably, there are thresholds þ1, þ2, þ3 (to be recovered in the course of scaling), such that (i,j) and (j,k) are ranked with a confidence rating of 3 if $|\Delta_{ij,jk}| >$ þ3, while a confidence rating of 2 implies that þ3 $> |\Delta_{ij,jk}| >$ þ2, and so on. In this interpretation, a tie is the special case of a zero-confidence judgment. However, I am loath to digress any further into dissimilarity-difference scales. It is hard enough to assign scale values to the dissimilarities themselves: if it were any easier, the triadic form of data would not be necessary.

The Gestalt-fusion paradigm: a special case

A modified version of PC-Triskele was written to replicate a series of explorations of colour space by Stalmeier and de Weert [1991a, 1994]. The stimuli are presented in the 'Star-of-David' form (described in Chapter 3), and the subject indicates whether i or k is closer to j by pressing a left- or right-pointing arrow key, according to the direction of the triangle formed by gestalt fusion of three coloured triangles with a central hexagon. The other two comparisons in $<i,j,k>$ are made separately (Complete Method of Triads).

During pilot studies, an interesting effect appeared: this is a transparency illusion. For some combinations of colours *i,j,k,* the eye tends to interpret the star-of-David in a different way, as a pair of overlapping triangles, the one more distant from the viewer being partially visible through the closer triangle, with the resulting mixture of colours being the central hexagon. This interpretation interferes with the left / right judgment, but it has the potential to provide useful information in its own right.

When the background is black, as in the Stalmeier-de Weert studies, the transparency illusion is most likely to appear when the central hexagon is brighter than its surrounding triangles. This becomes apparent from examining one of the data sets which Stalmeier generously provided, one not previously discussed, where $N = 13$ and the colours varied in luminosity as well as hue and saturation. The confounding effects of transparency make it difficult to reconstruct the known configuration from these data. A spurious dimension intrudes, in which the extremes of luminosity are similar, and a four-dimensional solution is required to recapture the known three-dimensional structure. However, the picture is clarified by ridding the data of all transparency-prone comparisons in which j is brighter than i or k; three dimensions then suffice.

In this situation with a black background, the coloured polygons are perceived as self-luminous. Let x_i, x_j, x_k be the points corresponding the the stimuli in some colour space. Then if

$$x_j = x_k + \beta_k x_i \qquad (0 \leq \beta \leq 1) \qquad (8.5a)$$

the display invites interpretation as a triangle of colour k and transparency β_k, overlapping a second triangle of colour i. Conversely, if

$$x_j = x_i + \beta_i x_k \qquad (0 \leq \gamma \leq 1) \qquad (8.5b)$$

then the i-coloured triangle is overlapping the k-coloured one, and has transparency β_i. In either case, j must be brighter than either of the colours contributing to it.

Suppose that we give subjects the options of indicating that i,j,k evince the transparency illusion. Suppose, further, that a subject sees a triad as a k-coloured triangle overlapping an i-

coloured one, given x_i, x_j, x_k which have no exact solution for β_k in (8.5a). Then this is vector information, comparable to the forms of data considered at the end of Chapter 6. Each 'transparency' response is an indication that x_i, x_j, x_k should be adjusted to make (8.5a) soluble.

The argument is speculative, since the colour-triad project is in abeyance until enhancements to a PC monitor can be acquired for displaying specified colours with sufficient accuracy and stability.

Mac-Triskele

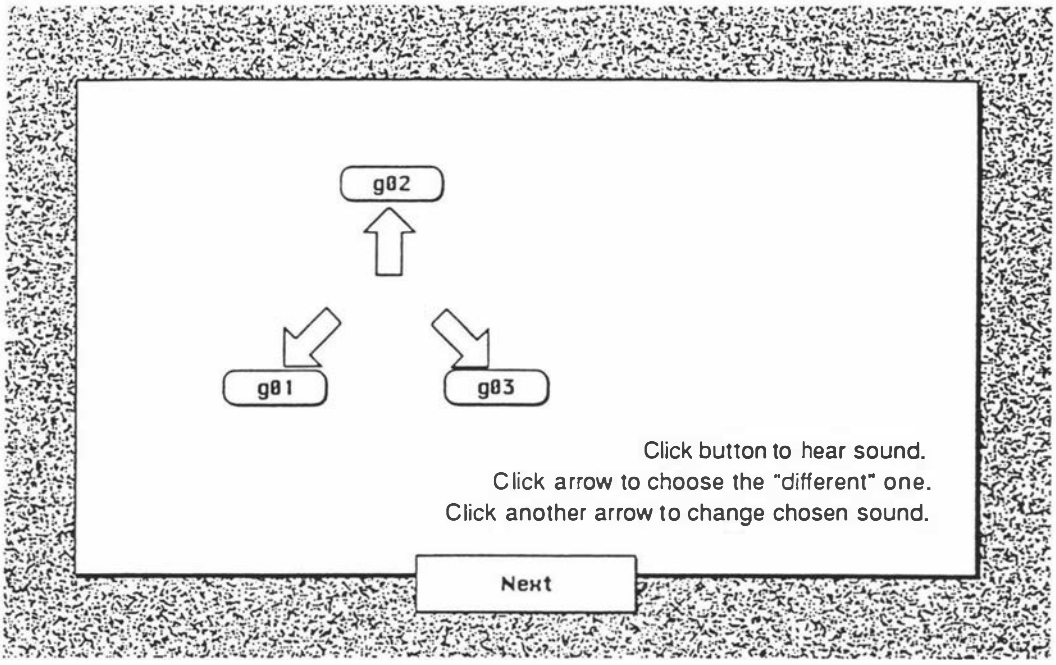
The differences between the PC and the Max platforms are substantial. On the latter, the user interface is paramount; the computational sections of a program are relegated to secondary status. For the Mac version, I reorganised the structure of Triskele to demarcate user-interaction from the processing of responses.

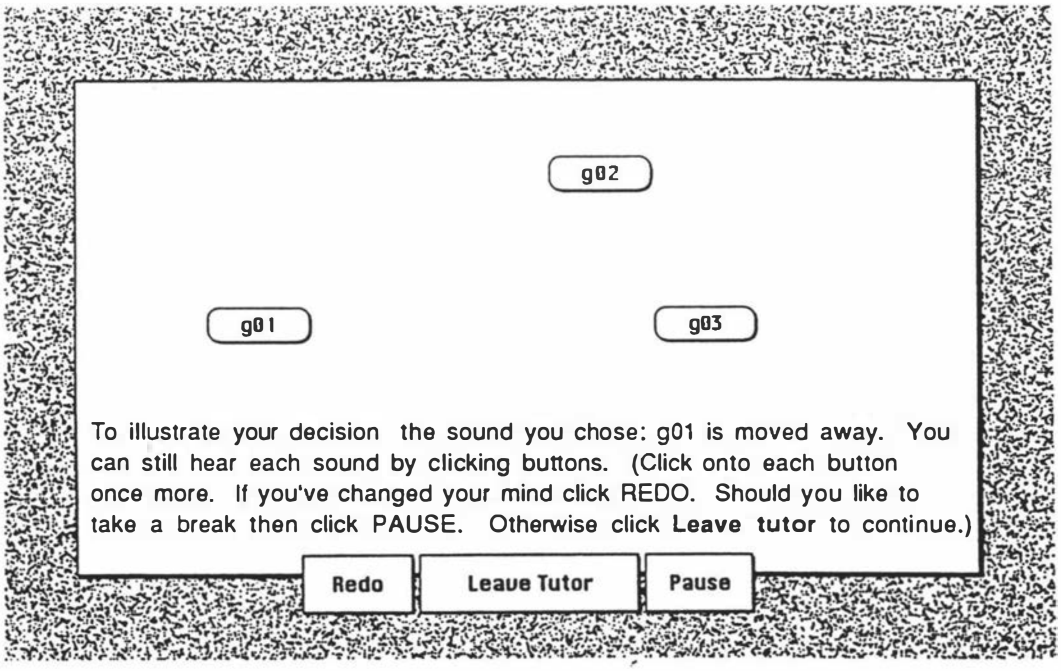
A Mac consultant (Steve Paris, of Wellington) converted the processing half into Apple Pascal. In its compiled form, as "executable external routines", this was inserted into the interaction half, which takes the form of a "stack" prepared in the Hypercard scripting language (the programmer being Mario Leonti, a Mac specialist in Palmerston North).

The stack contains screen images, "cards", which are repeated for each triad. They contain buttons which the subject clocks on with the mouse to play the sounds and other buttons with which the subject records a decision. Other screens explain the task, and offer tuition and help (figure 8.3). Further details appear in Kirkland, Bimler and Leonti [1992, 1993].

Mac-Triskele is readily modified for different sensory modalities, but the current version is sound-centred, using the Mac's 'Sound Library' and 'Sound Librarian' facilities. Sounds are fed in, sampled at 22 kHz, and stored. The Sound Librarian allows the experimenter to select the sounds for a Triskele session by clicking on buttons with the mouse, the order of selection also being the order of presentation.

Figure 8.3 Sample screens from Mac-Triskele session





Canine Heartbeats

Members of the Massey University Veterinary Faculty have used Triskele intensively, the stimuli being 20 tape-recordings of canine heartbeats exhibiting various forms of heart abnormality. The tapes are a standard training set. For this application the stored sounds were looped, so that they keep playing for as long as the mouse button is held down: the unit of repetition being sometimes a single heartbeat cycle, sometimes several cycles, since some abnormalities are characterised by their intermittency.

The attractive feature of heartbeats, from psychology's point of view, is that a trained, experienced listener can diagnose a particular syndrome and its degree of severity from the sound alone. However, isolating the acoustic properties giving rise to a diagnosis can be very difficult. In this, heartbeats are akin to baby cries, and numerous other complex sounds, including many used by clinicians.

As well as the qualities, whatever they are, which make diagnosis possible, a sound has other, accidental, irrelevant features, which the expert listener has learned to ignore. The relevant features are not necessarily to be localised to a particular spectral band or a particular segment of the time domain; they may take the form of complex relationships between several parts of the spectrum or time domain, allowing them to survive transformations (such as absorption of frequency bands, by intervening walls in the case of cries, by varying thicknesses of chest, for heartbeats).

The chief objective of this project was to map an idealised (average) experienced clinician's mental representation of heart abnormalities. The dimensionality of 'heartbeat space' was not initially known. Indeed, it is not certain whether a spatial model or some kind of tree structure is more appropriate. 20 stimuli, hopefully, include enough specimens of the various syndromes, in enough levels of severity, to reveal the relationships between them.

A secondary objective is to explore how actual subject's mental maps vary with their degree of proficiency. Are all novice listeners the same, with a standard progression of intermediate maps as they accumulate experience at making the important distinction? Or are there as many mental maps as there are novices? Landis, Silver, Jones & Messick [1967] explored similar questions for visual displays of airport flight-control situations. See also Michon's study [1972], where the stimuli were complex sounds, and triadic data from trainees were converted into dendrograms and compared against dendrograms from experienced listeners.

The research I am describing was conducted by three Massey students – Emma Barraclough, Greg Jones and Don Kirkland. With Triskele still in a pilot-study stage, 20 stimuli were thought to be too many for any one subject, so this was reduced to 14 stimuli per Triskele session, with the 14-stimulus subsets containing all 20 between them. The experimenters varied the stimuli in each 14, and the order of presentation, according to no particular system. In each session, 9 stimuli formed the phase 1 basis set (84 triads), and locating each remaining stimulus took 14 triads (6 in phase 2a, 8 in 2b), for a total of 154 triads. Primary comparisons only were elicited (odd-one-out triads).

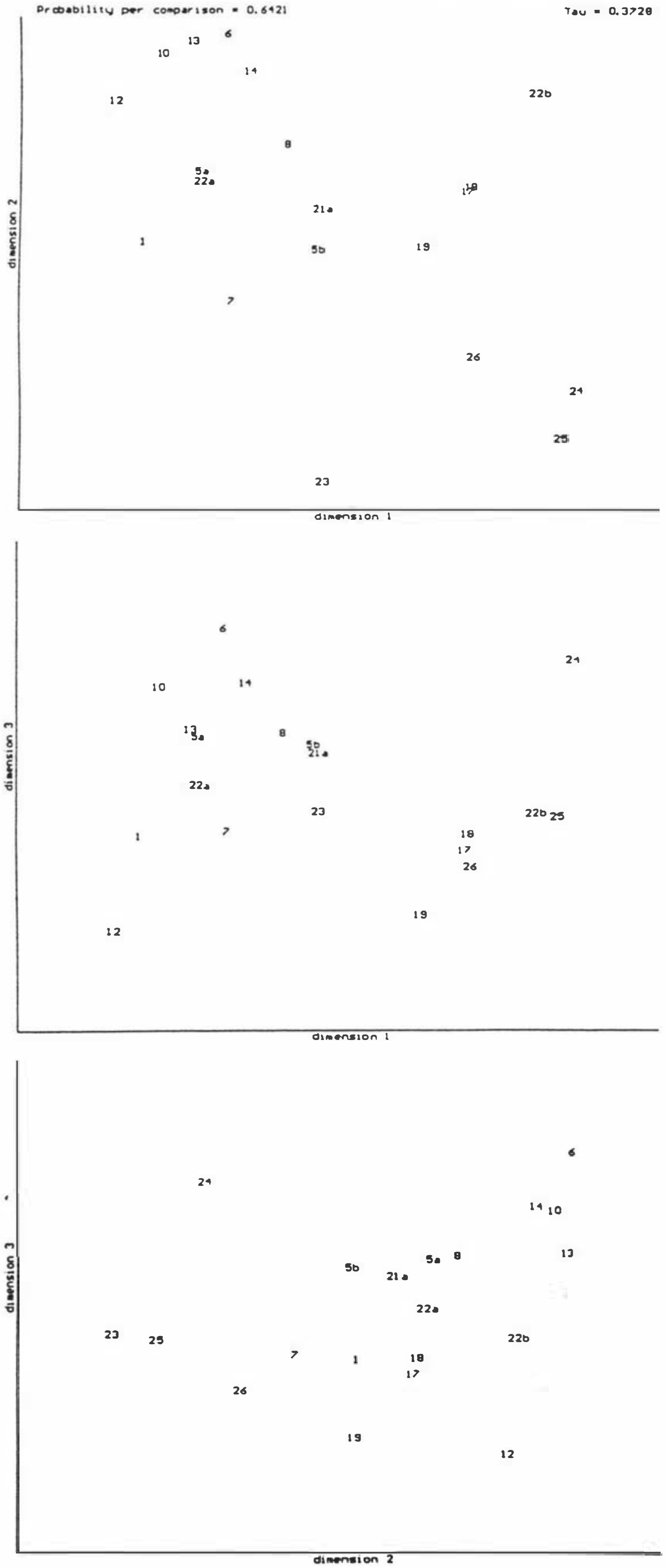
The database consists of triad sets from 18 Triskele sessions. Four are from Professor Boyd Jones, who also completed half of a fifth for a colleague. Several staff members sat through two sessions.

I have been provided with identifications for 14 of the stimuli, as follows:

1: normal dog, mitral region
5a: systolic murmur
5b: systolic murmur
6: musical systolic murmur
8: systolic murmur of pulmonic stenosis
10: systolic crescendo
12: diastolic murmur
13: machinery murmur
14: systolic murmur, mitral region
18: systolic click, mitral
19: diastolic gallop, mitral
23: incomplete dropped beat
24: atrial fibrillation
26: premature ventricular beat.

With these labels in hand, it becomes possible to reach some conclusions from the map produced by combining all the data sets (figure 8.4). The most noticeable feature is the

Figure 8.4 Three-dimensional configuration for triadic data for 20 canine heartbeats

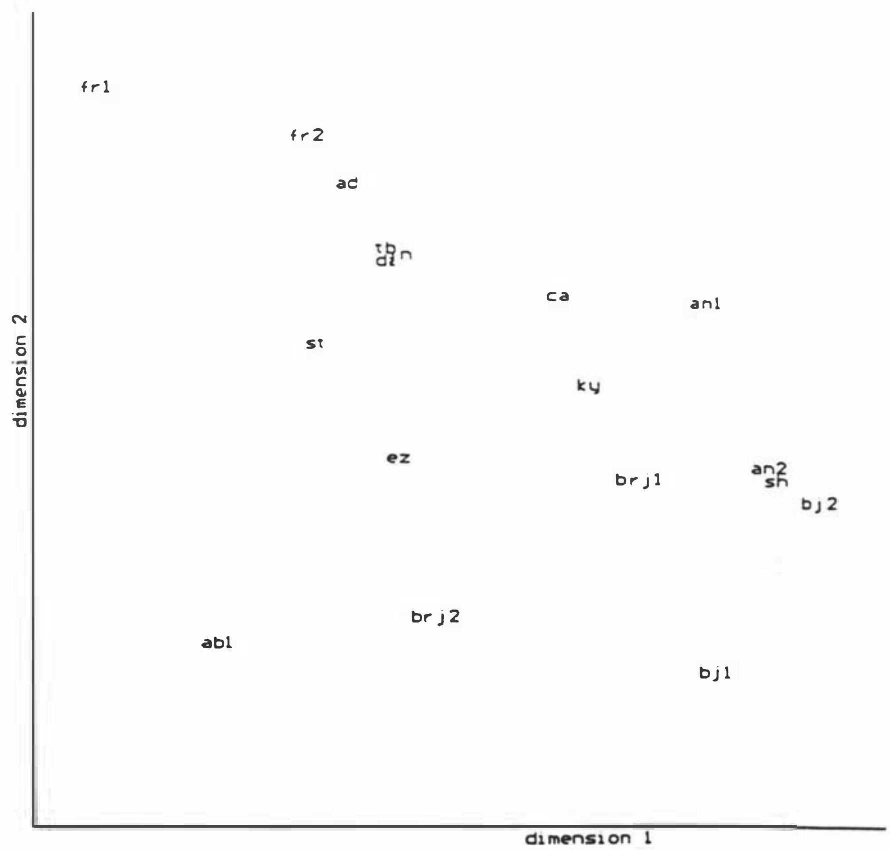


separation of 'murmur' sounds from others along the first dimension. The other two dimensions are harder to identify, but they seem to be significant to the diagnosticians. The INDSCAL model was used to rotate figure 8.4 to the most relevant orientation.

It seems from the subject space produced by the INDSCAL option (figure 8.5) that a major form of variation between our informants is the importance they attach to the first dimension. They ranged from 'an2' at one end, and 'bj1' and 'bj2' (both aliasses for Prof. Jones) to 'fr1', 'fr2' and 'ad' at the other. It is worth noting that the Points-of-View analysis reveals much the same picture: variation along a one-dimensional range, with the extremes (the viewpoints in their purest form) represented by 'ad' and 'an2'.

Given the elongation of the configuration along the first dimension, it is possible that the second and third dimensions each have more than one rôle, i.e. that the quality separating beats 2 and 6 to the left of the configuration is different from that separating 24 and 25 to the right, although both are accommodated within a third dimension.

Figure 8.5 Heartbeat triad data: Subject space from INDSCAL for 18 subjects



Feedback from the triad contributors was favourable: they found the configuration meaningful. Nevertheless, the question was whether diagnostic terms could be embedded in the perceptual space derived from triads. Do the two forms of element actually occupy a coomon mental representation? Are the forms of auditory variation contributing to "difference", according to the expert listeners (not necessarily obvious forms of difference to a spectrograph or untrained listener) the ones used to discriminate diagnoses and prognoses?

Cross-modal supplementary data

More data were collected to supplement the triads. This involved the 14 stimuli listed earlier, and 9 labels or diagnostic categories (possibly overlapping):

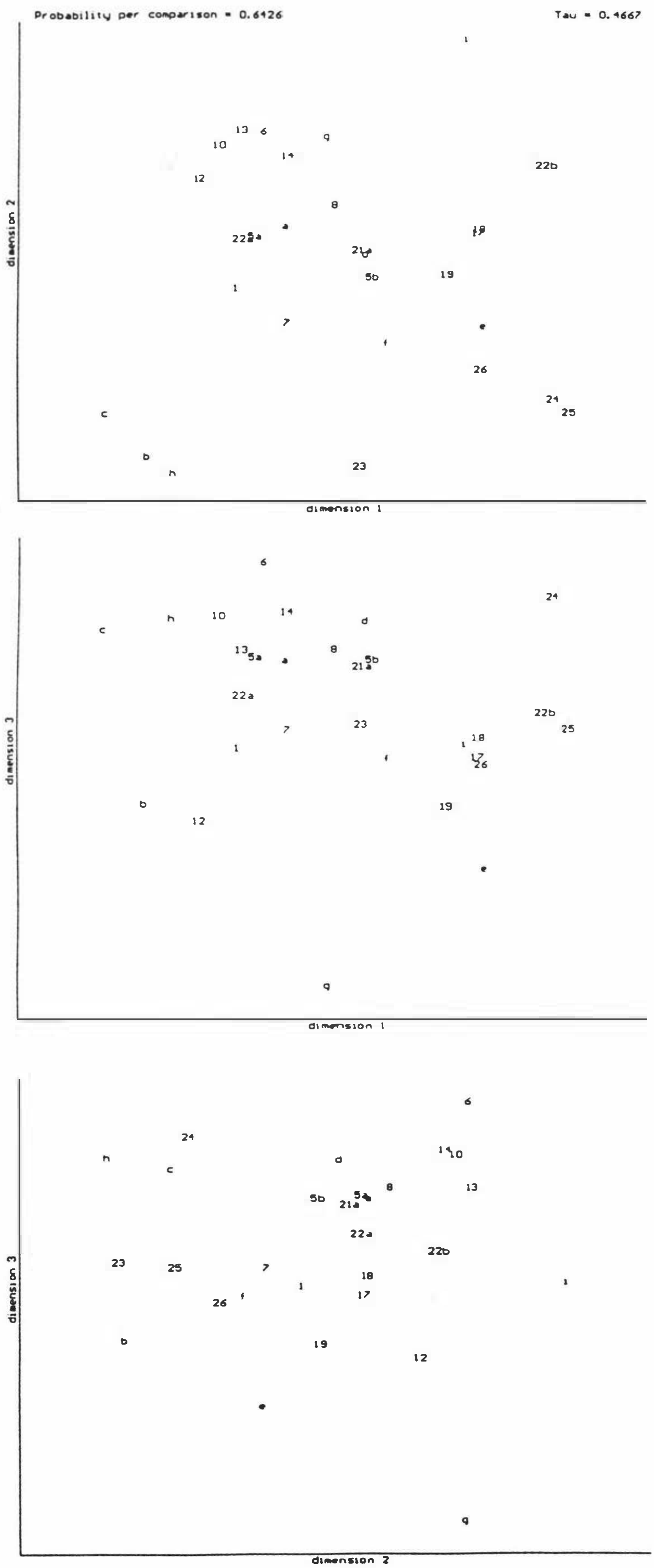
a Mitral inefficiency
b pulmonial stenosis
c tricuspid inadequacy
d ventricular septal defects
e split heart sounds
f patent ductus arteriosis
g aortic insuff iciency
h teratology of Follot
i aortic stenosis

For each stimulus, judges (six of them) picked the first, second and third most appropriate categories, making this rank 3/N cross-modal data. For some judges, some of the sounds were so clear-cut in diagnostic terms that only a "most appropriate" choice was possible.

In fact, scaling these data jointly with the triads did embed the categories in the same representation (see figure 8.6). There are some surprising features. Category b, 'pulmonic stenosis', is not as close to heartbeat 8 ('systolic murmur of pulmonic stenosis') as one might expect.

Further explorations are continuing, using scales to identify the axes. The scales are the low-resolution form discussed in Chapter 3; the subject (Professor Jones) listens to each sound in turn, with a discriminating feature in mind, deciding whether or not the sound exhibits it.

Figure 8.6 20 canine heartbeat stimuli, plus 9 diagnostic labels, 'a' to 'i'



Implications for Training

I want to make a few points about the educational possibilities of multidimensional scaling methods. I will avoid details since the topic is not central to this dissertation, but it is one of the reasons for the Veterinary Faculty's enthusiasm, and reveals another aspect of the Method of Triads.

The problem facing a vet. student is to learn to recognise auditory features (or absence of features) that can be used to situate a heart sound at its correct position in heartbeat space, and to learn which features are irrelevant, signifying nothing by their presence or absence. One way to learn this is to listen to sounds in groups of three, two sharing a feature which is lacking in the third: in a word, triads. To borrow Kelly's terms [1955], a triad invites the listener to distinguish a construct from a contrast. A series of triads might begin with ones forming quite elongated triangles in heartbeat space, so that the odd-one-out is easily spotted, progressing to closer and closer approximations of equilateral triangles, with the difficulty of distinguishing the odd-one-out keeping pace with the student's growing skill. Each elongated triangle should be aligned with one of the discriminatory directions (not necessarily axes) identified through studies with scales.

Triads are not the only possibility – having a map of the space makes several options available. One is to present the student with groups of four stimuli, with one in each group standing out [Isaac, 1970]. A third option again uses four stimuli at a time, the subject this time being to arrange each tetrad into closest pairs (i.e. to pick i, j, k, l so that $d_{ik} + d_{jl}$ and $d_{il} + d_{jk}$ are both greater than $d_{ij} + d_{kl}$).

I noted in passing in Chapter 4 that in situations where perceptual space has been mapped, triadic comparisons can be selected specifically to estimate a subject's dimensional weights. The selection can be interactive. A process of successive refinements similar to Triskele is possible. Given an estimated weights vector, i.e. an estimate of the private stimulus space for the subject, one can calculate how much information about the weights each triad can be expected to provide. Particularly informative triadic comparisons are ones between (i,j) and

(j,k) that are roughly aligned with different axes while $d_{m,ij}$ and $d_{m,jk}$ are roughly equal (similar statements can be made for other models fo individual variation).

Consider now the question of determining τ for a particular subject, again given a map of perceptual space. One approach is to start by guessing that τ is small, and to present comparisons between large and small distances (elongated triangles), every correct response being an indication that τ can be increased (as in Chapter 5) and that the elongation of the next triad can be less.

As well as mapping heartbeat space for sheer curiosity, we are concerned with the diagnostic possibilities. Locating a new stimulus within a well-landmarked configuration is equivalent to specifying the syndrome with the minimum amount of effort. Perhaps an atlas is a better analogy than a map. The first comparisons involving a novel stimulus specify the page; from there, successive comparisons straiten the scope. The goal is an optimum "search tree" (is this what experienced clinicians have?)

A second example is cries (at last). Here, there are the same dual concerns with diagnosis (considerable effort has gone into the possibility that some cries indicate forms of illness or stress or birth defect) and with distinguishing experienced from novice judges, so that the latter can receive feedback and targeted tuition.

9 CONCLUSIONS AND FUTURE DIRECTIONS

We have made little headway toward the original goal of preparing a comprehensive atlas of cry space, but have covered a lot of ground and forged tools of more general application in the process.

Concentrating on triadic and sorting data, this thesis has discussed sundry forms of data where the common feature is that they can be interpreted as greater than / less than comparisons between inter-point distances in an abstract perceptual space.

The advantages of this interpretation are several. It allows the introduction of Maximum Likelihood Estimation methods, as a straightforward extension of the familiar Stress-minimising method. Moreover, one can perform multi-dimensional scaling directly on the comparisons, as proposed by Johnson [1973], without converting them first into a matrix of dissimilarities; the degradation and distortions brought about by that Procrustean conversion are avoidable.

I have not attempted to optimise the details of this scaling. For didactic purposes and for ease of programming, the current version of MTRIAD adheres to a simple hill-descent algorithm for arranging points in perceptual space so as to maximise agreement between the modelled distance and the actual point, although more efficient algorithms exist. If switching to one of the alternative algorithms (e.g. the conjugate gradient method for estimating step size, or majorisation [de Leeuw, 1988] or other second-order hill-descent algorithms) we would seek to preserve the flexibility of hill-descent: the freedom to vary such things as the distance function, and the definition of the agreement being maximised.

In this dissertation I have not explored the interesting problem of representing perceptual / semantic structure in non-spatial or hybrid ways (e.g. trees). The question of choosing a tree which best describes a set of distance observations is the opposite of the problem considered in Chapter 7, of deriving the distances given a set of observed trees. I suspect that existing algorithms for fitting non-spatial representations to dissimilarity matrices, by iteratively minimising mismatches [de Soete, de Sarbo & Carroll, 1985], could be modified to accept dissimilarity *comparisons* (triads, etc.) as their raw material. A further possibility there is refining a tree model by acquiring comparisons interactively: an extension of the Triskele approach.

The distance comparison interpretation of data has corollaries for the question of efficiently selecting incomplete data, when human frailty precludes the collection of a complete set. I have looked at the opportunities for selecting and acquiring incomplete data interactively. An interesting special case arises when an unknown element is be positioned within a known configuration (for purposes of identification and perhaps diagnosis) with the fewest number of comparisons.

Related to this are conclusions one can draw about *which* comparisons are contained in particular data formats. When the sorting or hierarchical sorting or pick any/*N* methods are used as a way of scaling large stimulus sets, their lack of long-distance comparisons leaves the global structure of the configuration in doubt, unless they are complemented by other forms of data. Triads or dissimilarity rankings on a subset of stimuli can "cross-brace" the configuration. Other sources of global structure are preferences or scale ratings, which are difficult raw material for MDS on their own, fraught with potential artifacts because of the lack of direct constraints on the distances between ideal points or scale endpoints. Even analogy tasks are potentially a way of removing the unwanted flexibility from the configuration.

One goal for future research is to acquire analogy data as a test of the validity of the I-FEEL and Lightfoot expression solutions. In the description given in Chapter 6, the analogy task has the drawback that when it is used to test whether a spatial configuration of stimuli is an adequate model for people's mental representations, it incites subjects, by its wording, to think in the geometrical terms which one wishes to test. Fortunately the nature of the I-FEEL and Lightfoot stimuli make a non-geometrical form of analogy test possible.

This form involves a list of emotion-altering events (in the I-FEEL case, the faces are those of infants and young children, so this list might include parental interventions, examples being "loses toy", "familiar face", "loud noise", "change of nappy", etc.). Given one of these events, the task is to imagine it applied to a given facial expression, and to choose (from a small list of possible answers) the stimulus best representing the subsequent expression / emotion. A data set comprised of such choices would make it possible to tell whether each action corresponds to a vector in emotion space (x units increment along one axis, y units along another). For a valid spatial model of expressions, the vectors should be independent of position, with a given event producing the same displacement, whatever stimulus it is applied to.

I have demonstrated the practicality of pooling the responses of several subjects to amass enough data for recovery of a group configuration, without obscuring the variations between the subjects' personal versions of that configuration. Useful models of individual variation (e.g. the INDSCAL, IDIOSCAL, and Points-of-View models) allow solutions to be obtained for the sparse individual data-sets by imposing constraints, confining them to a low-dimensional subspace of configuration space (to a single point, if they are treated as replications). The appearance of gaining something for nothing is misleading.

The forms of data most useful for scaling large numbers of stimuli are low in redundancy, and *must* be pooled. I described a program (Triskele) which implements an interactive approach to lowering the redundancy and increasing the usefulness of triadic data, and considered the diagnostic implications of locating a novel stimulus, interactively, within a pre-existing perceptual map.

To date, the chief application of Triskele has been to map a 'heartbeat space' of canine heartbeat irregularities. Triskele has now been appraised with sufficient thoroughness to consider applying it to map a large set of cry stimuli.

One theme running through this thesis has been this problem of mapping large sets of stimuli ($N > 20$). As well as the Triskele program, which I hope is a useful contribution to the triadic method, I have presented a method for the analysis of sorting and hierarchical sorting data. The low demands made by these methods on subjects make them ideal for large-N data collection, so long as an adequate analysis for those data is available.

In some cases, the fine discrimination of subject variability is the objective, as considered in Chapters 4 and 8. There are several relevant lines of research.

Green, Jones and Gustafson [1987] used judged similarities between baby cries to map the perceptual space, and found differences in dimensional weights between sub-populations. One research priority is to look for the same kind of differences in triadic data. Are the variations discernible at the level of individuals, without averaging over groups? To be useful, such variations should be consistent enough to predict a subject's sex and parental status.

If the group configuration is known from other sources, we saw that some comparisons are more useful than others for locating a subject on the spectrum or spectra of individual variations. It becomes possible to acquire such comparisons selectively.

Analogous to the "selective deafness" of some populations [Green, Jones & Gustafson, 1987] – their insensitivity or lack of attunement to affordances in the cries which other subjects picked up – are the several forms of colourblindness.

Recall that as well as revealing the expected colour circle, ranked dissimilarities for a set of 9 colours provided a crude test of colour vision. The H-sorts of the 9-colour set and of the 16 colours of the D-15 stimulus set were not enough to recover the global structure of colour space (the configurations being locally one dimensional), but they contained enough information about subjects' dimensional weights, limited by the relatively minor parts these weights played in the comparisons.

"Informative" comparisons can be acquired more selectively using the various triadic methods: for instance, the Star-of-David experimental design. The Stalmeier-de Geert data sets provided a demonstration of this. I note that this process is an candidate for conversion into an interactive process, akin to Triskele. It comes down to hypothesising a set of dimensional weights for subjects and selecting triads for their potential to refine the hypothesis. The choice of triad having the greatest expected influence on the values assigned to the weights is affected by their provisional values.

This is one of my major conclusions. The dissimilarity comparisons implicit in several forms of data – triads, hierarchical sorting, and the D15 procedure – can define a subject's position in 'subject space', especially if the configuration of stimuli (the group space) is known. The IDIOSCAL model is particularly rewarding. Current tests for colour vision are designed to pick up blatant deviations from normal. They are not sensitive to variations within the bounds of 'normality', and leave many questions unanswered about colour-vision variations amongst the heterozygous female relatives of overtly colour-blind male homozygotes. Furthermore, knowing the precise extent of a subject's colour-vision deficiency can help with advising him what adaptations to make at home and at work. Work on fine-tuned triad tests is continuing.

Unlike colour-blindness, selective deafness is remediable through training [Green, Jones & Gustafson, 1987; but also the growth of sensitivity to dimensions of musical appreciation, Pollard-Gott, 1983; the effects of tuition in the recognition of complex sounds, Michon, 1972]. Thus, a potential application of the work reported here is to monitor the progress of a course of training, and to provide feedback for the trainee.

I reported research with canine heartbeats where many of the judges subjected themselves to repeated Triskele sessions, making it possible to plot their progress towards acquiring the perceptual space of an expert. Now that the "expert" configuration is known, future work with these stimuli will concentrate on determining how individuals differ from it, at stages in their training, using triads optimised for that purpose. We found that recognised syndromes (verbal labels) can be located in perceptual space too, so cross-modal forms of data can be used to build up the configuration and to probe to what degree a subject has internalised it.

The INDSCAL model is not the only way of accommodating individual variability. It may turn out that the extent of a person's colour-vision deficiency depends on the saturation of the colours[1],or that the development of expert listening involves greater changes in perceptual space than simply becoming attuned to a dimension of previously low salience. We must bear such possibilities in mind, and be prepared to use more complex models if the observations so dictate.

---

[1] Not to mention stimulus size, and intake of coffee – to mention just two parameters known to make a difference in borderline cases.

The line separating the effects of tuition from normal development is fine. Jacobowitz's conditional rank-order data [Young, 1975], obtained from children of different ages, show how the semantic structures of colours, body parts and kinship terms evolve. Miller and Gelman [1983] obtained triadic data to show the development of the concept of number.

The same analysis applies. I note the possibility that such data, analysed in the framework of INDSCAL or some other model of individual variation, may not need to be averaged beforehand over subjects presumed to be at similar developmental levels. Are a subject's dimensional weights, in the INDSCAL model, or a_m values in the Points-of-View model – his or her location in the trend of the combined data sets – a predictor of development?

The strategies and variant methods I have discussed for minimising the data requirements of MDS come to the fore when children are included among the subjects.

The emphasis in this thesis has been on developing methods, as opposed to experimental work. Consequently, none of the experimental conclusions made along the way are particularly novel or in contradiction to any long-held scientific consensus. This is just as well, since I have applied these methods in a number of fields, and relied upon their agreement with the consensus view in order to validate them.

APPENDIX W. WOODBLOCKS

An on-going market-research research project, aimed at determining which aspects of Paulownia wood contributed most to its desirability, was the source of the data analysed here. The researcher was Emma Barraclough, of Massey University, in association with Bruce Glass of the Forest Research Institute. Here I ignore the market-research aspects of the data and consider the raw material of their research – finished slabs of Paulownia wood, 20 x 85 x 285 mm, cut at a range of angles – as stimuli to be scaled, with their dimensionality to be determined. Appropriate sections of Chapters 6 and 7 discussed details of the analysis of the various forms of data collected.

One set of data consisted of 37 preference scales, for a basis set of 12 blocks. These were analysed in three dimensions with the Ideal-point model for preferences, the first two dimensions being shown as Figure W.1. Ideal points themselves are represented as '.' to distinguish them from the stimuli. Most of the variation in the third dimension involved the ideal points, rather than element points, an arrangement which is probably artifactual.

Figure W.1 Two of three dimensions obtained from preferences for 12 woodblocks (M = 37 ideal points, shown as full stops)

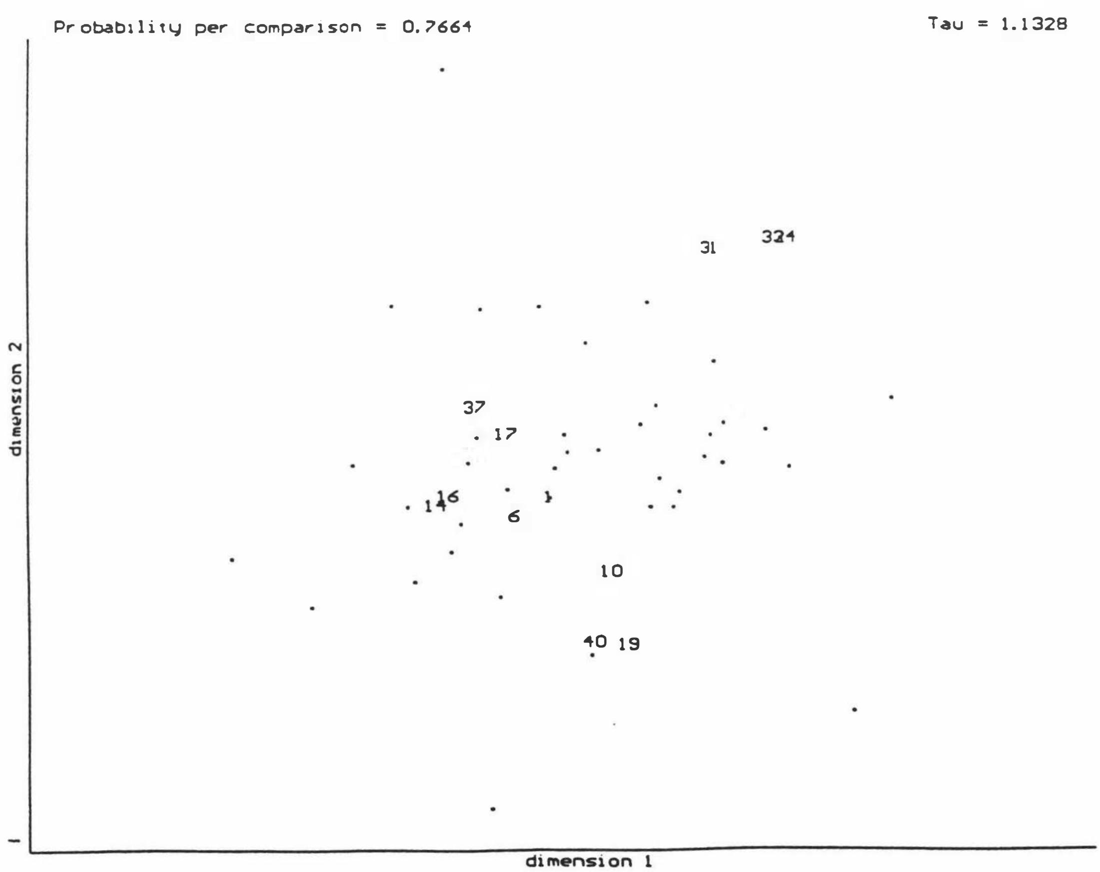


Figure W.1 provides a good basis for comparison with dissimilarity data. The same subjects provided hierarchical sorting data, using several procedures (synthetical H-sorting; analytical H-sorting; the analytical procedure, using photographs rather than the blocks themselves). Scaled with the reconstructed-dyad method, these result in a three-dimensional configuration. The third dimension manifests in subsets of the data, and appears to be both robust and interpretable. The overall configuration is similar to that derived from preferences. See figure W.3. Figure W.2 is a vote-counted configuration for the H-sorts; the same structure can be discerned in it, but much of the detail has been obscured by elements clumping together.

Figure W.2 Vote-count configuration, for 12 woodblocks, from hierarchical sorting data

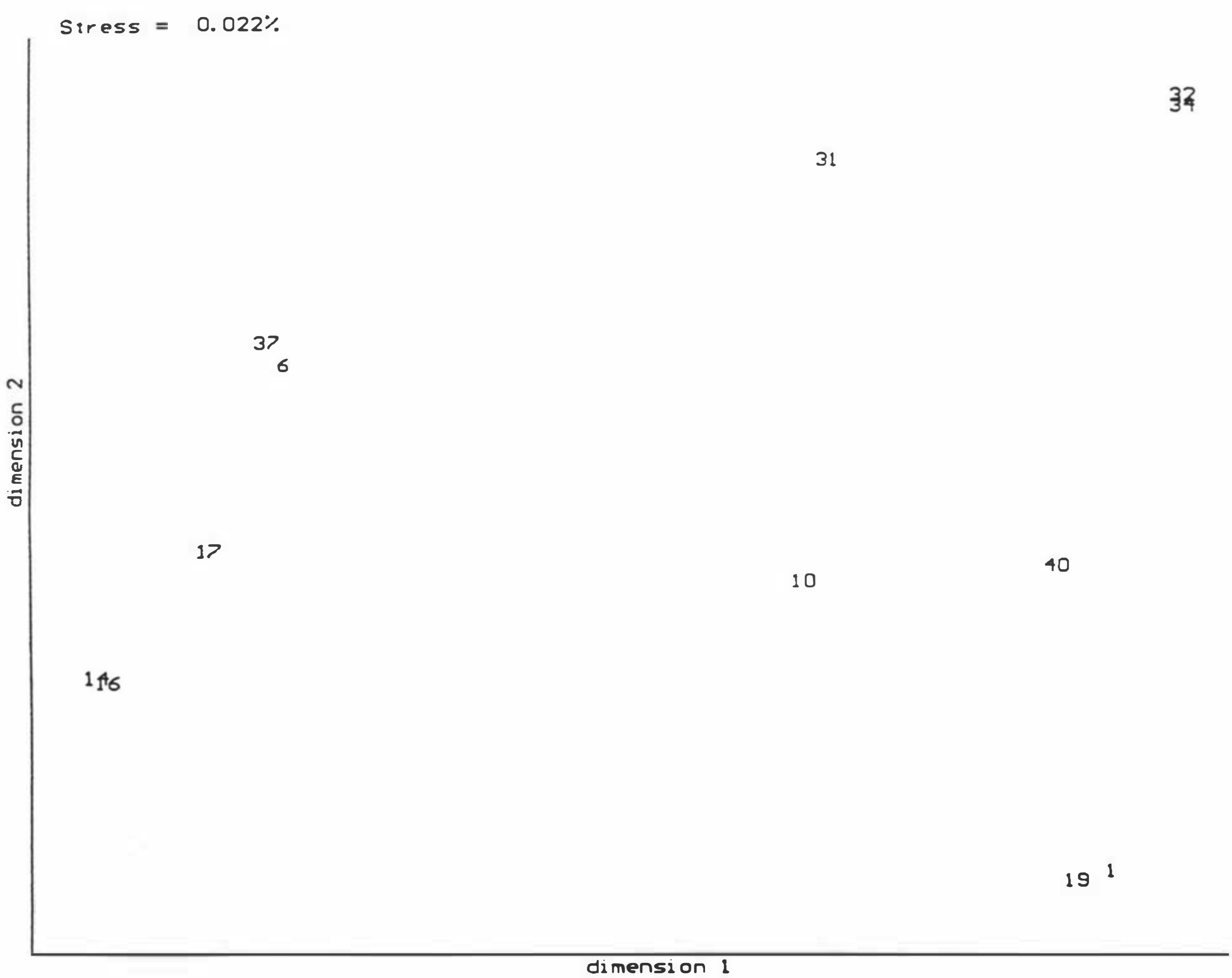
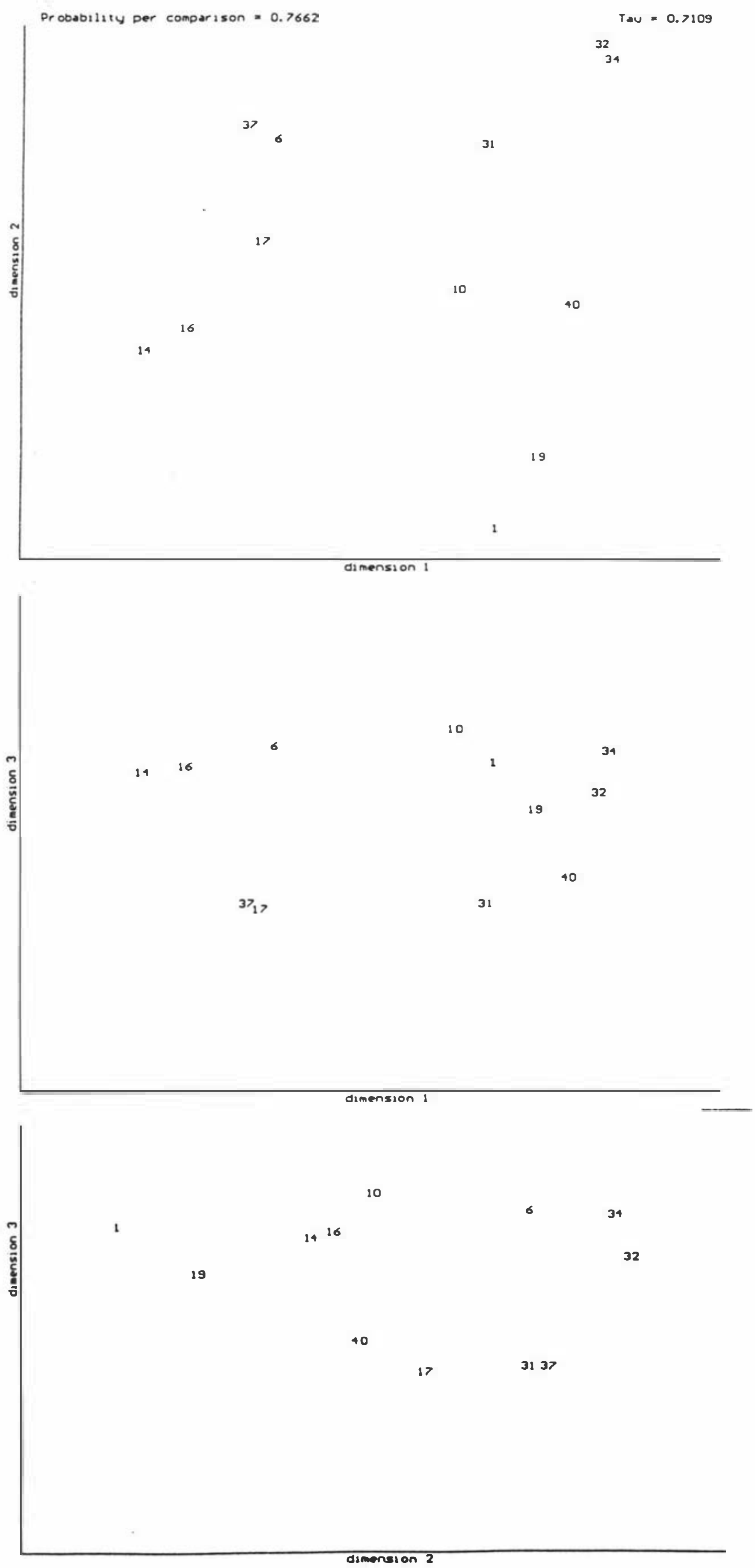


Figure W.3 Three-dimensional configuration for 12 woodblocks, using dissimilarity data (H-sorts)



Glass obtained ratings, on 18 scales, for a larger set of 24 blocks. When these scales are analysed, using the vector model, in two dimensions, the result is figure W.4. Again, the same overall structure appears. Because figure W.4 contains 12 elements which were not present in the previous dissimilarity data, it provides a test of the validity of the spatial scaling model. Barraclough proceeded to collect dissimilarity data (H-sorts, F-sorts, and preference rankings) for the complete 24-block set. When all these diverse forms of data are pooled, in combination with the scales, the result is figure W.5.

The accuracy with which the scales on their own predicted the positions of the 12 new stimuli gives some assurance that this research is on the right track.

Figure W.4 Configuration for 24 elements, derived from ratings on 18 scales (analysed in three dimensions)

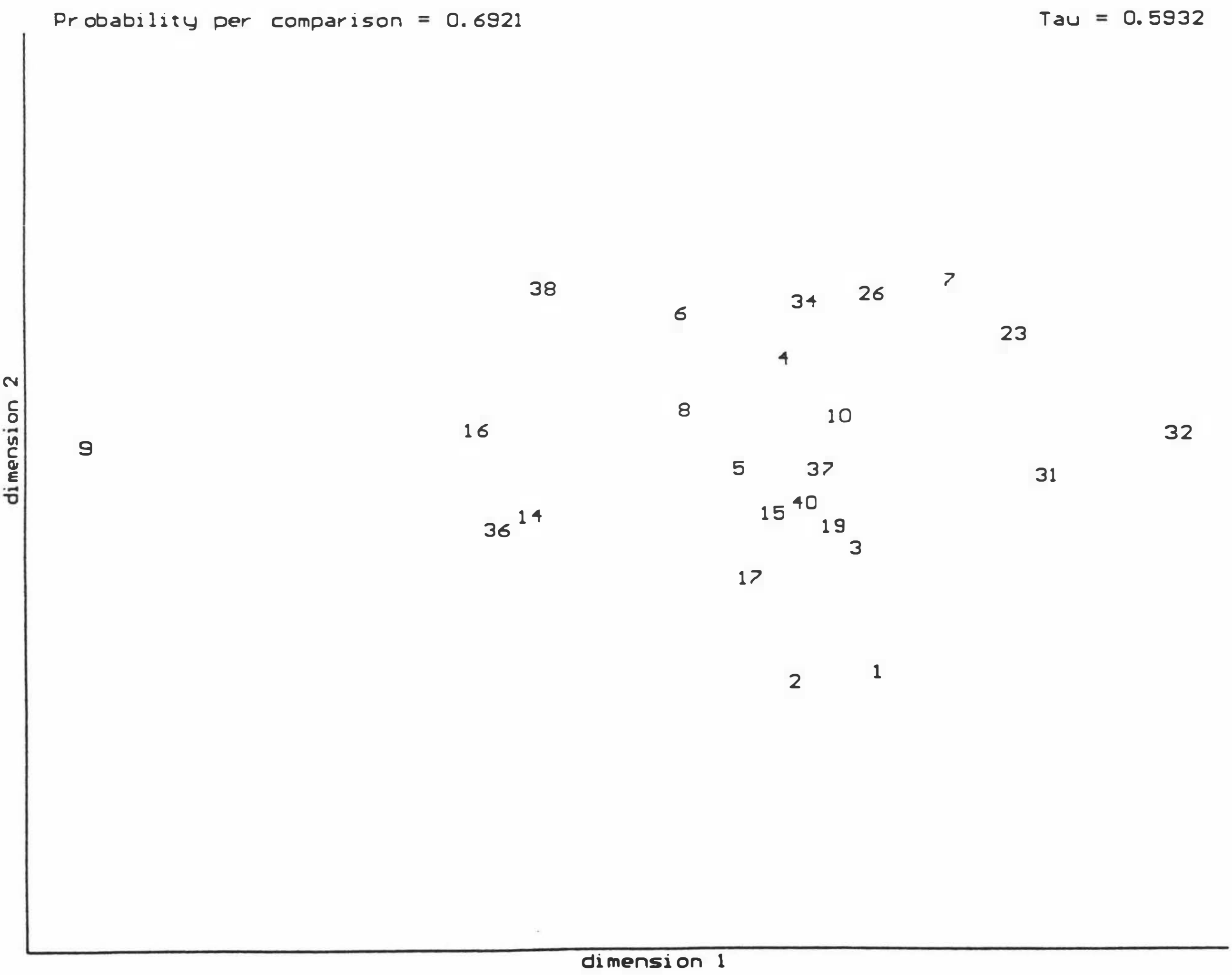
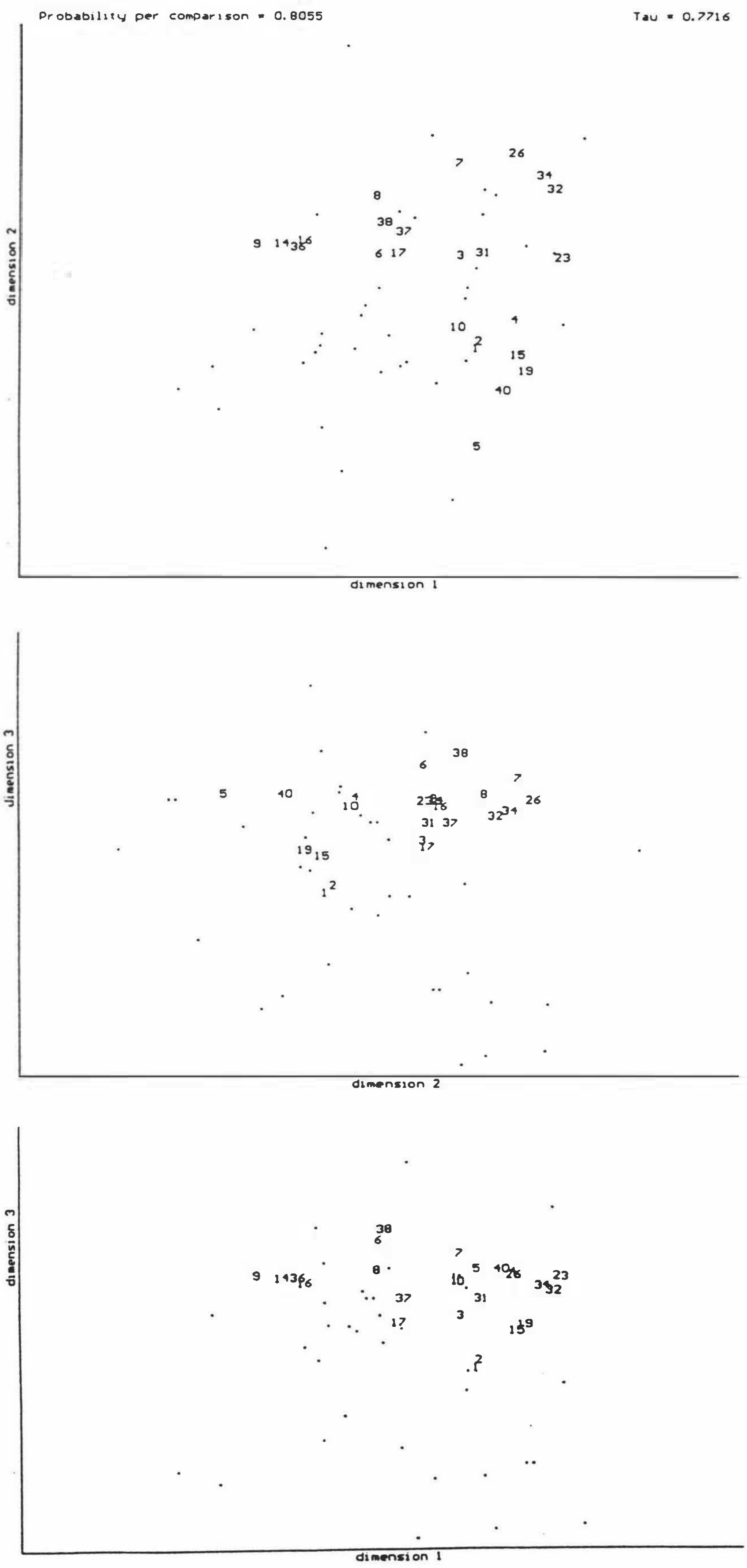


Figure W.5 Overall configuration for 24 woodblocks, combining preferences, H-sorts, F-sorts and scale-rating data



Bibliography

Abelson, R.P and Sermat, V. (1962) Multidimensional scaling of facial expressions. Journal of Experimental Psychology, 63(6), 546-554

Andrews, T.G. and Ray W. S. (1957) Multidimensional psychophysics: A method for perceptual analysis. Journal of Psychology, 44, 133-144

Anglin, J. M. (1970) The Growth of Word Meaning. Cambridge, Mass.: MIT Press

Arabie, P. (1991) Was Euclid an unnecessarily sophisticated psychologist? Psychometrika, 56, 567-587

Arabie, P. and Boorman, S. A. (1973) Multidimensional scaling of measures of distance between partitions. Journal of Mathematical Psychology, 10, 148-203

Arabie, P., Kosslyn, S. and Nelson, K. (1975) A multidimensional scaling study of visual memory of five-year-olds and adults. Journal of Experimental Child Psychology, 19, 327-345

Arnold, J. B. (1971) A multidimensional scaling study of semantic distance. Journal of Experimental Psychology, 90 (2), 349-372 (Monograph)

Attneave, F. (1950) Dimensions of similarity. American Journal of Psychology, 63, 516-556

Balzano, G. J. (1977) Chronometric studies of the musical interval sense. Unpublished Ph.D. thesis, Stanford University. University Microfilms order number 77-25,643

Bechtel, G. G. (1976) Multidimensional preference scaling (Methods and Models in the Social Sciences 6). The Hague: Mouton & Co.

Berlin, B., Breedlove, D. E., and Raven. P. H. (1968) Covert categories and folk taxonomies. American Anthropologist, 70 (2), 290-299

Bersted, C. T., Brown, B. R., and Evans, S. H. (1970) Evaluation of unconstrained sorting data. Behavior Research Methodology and Instrumentation, 2, 108-110

Bissett, R., and Schneider, B. (1992) Spatial and conjoint models based on pairwise comparisons of dissimilarities and combined effects: Complete and incomplete designs. Psychometrika, 56, 684-698

Boorman, S. A. and Arabie, P. (1972) Structural measures and the method of sorting. In: Multidimesional Scaling: Theory and applications in the Behavioral Sciences (eds. Shepard, Romney, Nerlove)

Boorman, S. A. and Olivier, D. C. (1973) Metrics on spaces of finite trees. Journal of Mathematical Psychology, 10 (1), 26-59

Borg, I. and Leutner, D. (1985) Measuring the similarity of multidimensional scaling configurations. Multivariate Behavioral Research, 20, 325-334

Borg, I. and Lingoes, J. (1987) Multidimensional similarity structure analysis. Springer-Verlag.

Borgatti, S. P. (1991) Provisional Documentation Manual for ANTHROPAC 2.6 (90 pages). Department of Sociology, University of South Carolina

Boynton, R. M. and Gordon, J. (1965) Bezold-Brücke hue shift measured by color-naming technique. Journal of the Optical Society of America, 55 (1), 78-86

Breiger, R. L., Boorman, S. A. and Arabie, P. (1975) An algorithm for clustering relational data, with applications to social network analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology, 12, 328-383.

Bricker, J. M. and Pruzansky, S. (1969) A comparison of sorting and pairwise similarity judgment techniques for scaling auditory stimuli. Paper 6E4, 78th Meeting, Acoustical Society of America.

Bums, B. and Shepp, B. E. (1988) Dimensional interactions and the structure of psychological space: The representation of hue, saturation and brightness. Perception and psychophysics, 43 (5), 494-507

Burton, M. L. (1972) Semantic dimensions of occupation names. In: Multidimensional Scaling: Theory and applications in the Behavioral Sciences, Volume II (eds. R. N. Shepard, A. K. Romney, S. B. Nerlove). NY: Seminar Press

Burton, M. L. (1975) Dissimilarity measures for unconstrained sorting data. Multivariate Behavioral Research, 10 (4), 409-423

Burton, M. L. and Nerlove, S. B. (1976) Balanced designs for triads tests: Two examples from English. Social Science Research, 5, 247-267

Burton, M. L. and Romney, A. K. A multidimensional representation of rôle terms. American Ethnologist, 2, 397-497

Carroll, J. D. (1972) Individual differences and multidimensional scaling. In: Multidimensional Scaling: Theory and applications in the Behavioral Sciences, Volume I (eds. R. N. Shepard, A. K. Romney, S. B. Nerlove). NY: Seminar Press

Carroll, J. D. (1976) Spatial, non-spatial and hybrid models for scaling. Psychometrika, 41, 439-463

Carroll, J. D. and Chang, J.-J. (1970) Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckart-Young' decomposition. Psychometrika, 35 (3), 283-319

Carroll, J. D. and Chang, J.-J. (1972) SIMULES: ***simu****l*taneous *l*inear *e*quation *s*caling. Proceedings, 80th Annual Conference, American Psychological Association, pp.11-12

Carroll, J. D. and Wish, M. (1974) Models and methods for three-way multidimensional scaling. In: Contemporary Developments in Mathematical Psychology, Volume II: Measurement, Psychophysics and Neural Information Processing (eds. A. H. Krantz, R. C. Atkinson, R. D. Luce, P. Suppes)

Church, A. T. and Katigbak, M. S. (1989) Internal, external and self-report structure of personality in a non-Western culture: An investigation of cross-language and cross-cultural generalizability. Journal of Personality and Social Psychology, 57 (5), 857-872

Clark, H. H. (1968) On the use and meaning of prepositions. Journal of Verbal Learning and Verbal Behavior, 7, 421-431

Clark, W. C., Ferrer-Brechner, T., Janal, M. N., Carroll, J. D. and Yang, J. C. (1989) The dimensions of pain: A multidimensional scaling study of cancer patients and healthy volunteers. Pain, 37, 23-32

Clark, W. C., Janal, M. N. and Carroll, J. D. (1989) Multidimensional pain requires multidimensional scaling. In: Issues in Pain Measurement (eds Chapman & Loeser). NY: Raven Press

Clarke, M. J. (1976) The identification of substitutes among selected fruits using multidimensional scaling. Market Research Centre, Massey University, Technical Report 6

Cliff, N., Giraud, R., Green, R. S., Kehoe, J. F. and Doherty, L. M. (1977) INTERSCAL: A TSO FORTRAN IV program for subject-computer interactive multidimensional scaling. Educational and Psychological Measurement, 37, 185-188

Cliff, N. and Young, F. W. (1968) On the relation between unidimensional judgments and multidimensional scaling. Organizational Behavior and Human Performance, 3, 269-285

Coombs, C. H. (1964) A Theory of Data. New York; Wiley

Coxon, A. P. M. and Jones, C. L. (1978) The images of occupational prestige.
— (1979a) Class and hierarchy: The social meaning of occupations.
— (1979b) Measurement and meanings: Techniques and methods of studying occupational cognition. London: Macmillan

Coxon *et al* (1975) Project on Occupational Cognition: Sociological Aspects of Subjective Occupational Structures [computer files]. Colchester: ESRC Data Archive

D'Andrade, R. G. (1978) U-statistic hierarchical clustering. Psychometrika, 43 (1), 59-67

Delbeke, L. (1968) Construction of preference spaces: An investigation into the applicability of multidimensional scaling models. Publications of the University of Louvain, Louvain-Leuven.

de Leeuw, J. (1988) Convergence of the majorization method for multidimensional scaling. Journal of Classification, 5, 163-180

de Soete, G., DeSarbo, W. S. and Carroll J. D. (1985) Optimal variable weighting for hierarchical clustering: An ALS algorithm. Journal of Classification, 2, 173-192

de Soete, G. and Winsberg, S. (1993) A Thurstonian pairwise choice model with univariate and multivariate spline transformations. Psychometrika, 58 (2), 233-256

Deutsch, D. (1992) Some new pitch paradoxes and their implications. Philosophical Transactions of the Royal Society of London B, 336, 391-397

Dong, H.-K. (1983) Method of complete triads: An investigation of unreliability in multidimensional perception of nations. Multivariate Behavioral Research, 18, 85-96

Drasgow, F. and Jones, L. E. (1979) Multidimensional scaling of derived dissimilarities. Multivariate Behavioral Research, 14, 227-244

Engen, T., Levy, N. and Schlosberg, H. (1958) The dimensional analysis of a new series of facial expressions. Journal of Experimental Psychology, 55, 454-458

Fillenbaum, S. and Rapoport, A. (1971) Structures in the subjective lexicon. NY: Academic Press

Fischer, W. and Micko, H. C. (1972) More about metrics of subjective spaces and attention distributions. Journal of Mathematical Psychology, 9, 36-54

Fowlkes, E. B. and Mallows, C. L. (1983) A method for comparing two hierarchical clusterings. Journal of the American Statistical Association, 78, 553-569 (plus discussion)

Friendly, M. L and Glucksberg, S. (1970) On the description of subcultural lexicons: A multidimensional approach. Journal of Personality and Social Psychology, 14 (1), 55-65

Garner, W. R. and Felfoldy, G. L. (1970) Integrability of stimulus dimensions in various types of information processing. Cognitive Psychology, 1, 225-241

Giraud, R. and Cliff, N. (1976) A Monte Carlo evaluation of interactive multidimensional scaling. Psychometrika, 41 (1), 43-64

Gladstones, W. H. (1962a) A multidimensional study of facial expression of emotion. Australian Journal of Psychology, 14 (2), 95-100

Gladstones, W. H. (1962b) Empirical comparisons of three multidimensional scaling methods. Australian Journal of Psychology, 14 (3), 197-201

Goldstone, R. (1994) An efficient method for obtaining similarity data. Behavioral Research Methods, Instruments and Computers, 26 (4), 381-386

Gordon, A. D., Jupp, P. E. and Byrne, R. W. (1989) The construction and assessment of mental maps. British Journal of Mathematical and Statistical Psychology, 42, 169-182

Graef, J. and Spence, I. (1979) Using distance information in the design of large multidimensional scaling experiments. Psychological Bulletin, 86 (1), 60-66

Green, J. A., Jones, L. E. and Gustafson, G. E. (1987) Perception of cries by parents and nonparents: Relation to cry acoustics. Developmental Psychology, 23 (3), 370-382

Green, P. E. and Rao, V. R. (1971) Multidimensional scaling and individual differences. Journal of Marketing Research, 8, 71-77

Green, R. S. and Bentler, P. M. (1979) Improving the efficiency and effectiveness of interactively selected MDS data designs. Psychometrika, 44 (1), 115-119

Gregson, R. A. M. (1966a) Theoretical and empirical multidimensional scalings of taste mixture matchings. British Journal of Mathematical and Statistical Psychology, 19 (1), 59-76

Gregson, R. A. M. (1966b) Cross-modal matching of histograms and four-component taste mixtures, with one component fixed, under two pacing conditions. Perceptual and Motor Skills, 23, 198-190

Gustafson, G. E.. and Green, J. A. (1989) On the importance of fundamental frequency and other acoustical features to cry perception and infant development. Child Development, 60, 772-780

Guttman, L. (1968) A general non-metric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33 (4), 469-506

Hall, J. L. and Schroeder, M. R. (1972) Monaural phase effects for two-tone signals. Journal of the Acoustical Society of America, 51, 1882-1884

Hamer, R. M. (1981) Nonmetric interactive multidimensional scaling with multiple subjects. Applied Psychological Measurement, 5 (3), 349-354

Hanson, G. (1967) Dimensions in speech sound perception: An experimental study of vowel perception. Ericsson Technics, 23 (1), 3-175

Harvey, L. O. Jnr. and Gervais, M. J. (1978) Visual texture perception and Fourier analysis. Perception and Psychophysics, 24 (6), 534-542

Harvey, L. O. Jnr. and Gervais, M. J. (1981) Internal representation of visual texture as the basis for the judgement of similarity. Journal of Experimental Psychology: Human Perception and Performance, 7 (4), 741-753

Helm, C. E. (1964) Multidimensional ratio scaling analysis of perceived color relations. Journal of the Optical Society of America, 54, 256-262

Helm, C. E. and Tucker, L. R. (1962) Individual differences in the structure of color perception. American Journal of Psychology, 75, 437-444

Henley, N. M. (1969) A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 8, 176-184

Henry, W. A. and Stumpf, R. V. (1975) Time and accuracy measures for alternative multidimensional scaling data collection methods. Journal of Marketing Research, 12. 165-170

Hojo, H. (1993) A new nonmetric multidimensional scaling method for sorting data. Japanese Psychological Research, 35 (3), 129-139

Hollins, M., Faldowski, R., Rao, S. and Young. F. W. (1993) Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. Perception and Psychophysics, 54 (6), 697-705

Hubert, L. J. and Arabie, P. (1985) Comparing partitions. Journal of Classification, 2, 193-218

Hubert, L. J. and Levin, J. R. (1976) Evaluating object set partitions: Free-sort analysis and some generalizations. Journal of Verbal Learning and Verbal Behavior, 15, 459-470

Hyman, R. and Well, A. (1968) Perceptual separability and spatial models. Perception and Psychophysics, 3, 161-165

Indow, T. (1968) Multidimensional mapping of visual space with real and simulated stars. Perception and Psychophysics, 3, 45-53

Indow, T. and Aoki, N. (1983) Multidimensional mapping of 178 Munsell colors. Color Research and Application, 8 (3), 145-152

Indow, T. and Kanazawa, K. (1960) Multidimensional mapping of Munsell colors varying in hue, chroma and value. Journal of Experimental Psychology, 59, 330-336

Indow, T. and Uchizono, T. (1960) Multidimensional mapping of Munsell colors varying in hue and chroma. Journal of Experimental Psychology, 59, 321-329

Isaac, P. D. (1970) Dissimilarities as indices of individual perceptual structure. Perception and Psychophysics, 7 (4), 229-233

Johnson, S. C. (1967) Hierarchical clustering schemes. Psychometrika, 32 (3), 241-254

Johnson, R. M. (1973) Pairwise nonmetric multidimensional scaling. Psychometrika, 38 (1), 11-18

Jones, R. A. and Ashmore, R. D. (1973) The structure of intergroup perception: Categories and dimensions in views of ethnic groups and adjectives used in stereotype research. Journal of Personality and Social Psychology, 25 (3), 428-438

Jones, R. A., Sensenig, J. and Ashmore, R. D. (1978) Systems of values and their multidimensional representations. Multivariate Behavioral Research, 13, 255-270

Junge, K. (1991) The geometric represntation of similarity and dissimilarity: Extension to measures of dispersion and comments on the distance model of subjective dissimilarity. Scandinavian Journal of Psychology, 32, 154-163

Kehoe, J. and Reynolds, T. J. (1977) Interactive multidimensional scaling of cognitive structure underlying person perception. Applied Psychological Measurement, 1 (2), 155-169

Kelly, G. A. (1955) The Psychology of Personal Constructs. NY: Norton

Killam, R. N., Lorton P. V. and Schubert, E. D. (1975) Interval recognition: Identification of harmonic and melodic intervals. Journal of Music Theory, (2), 212-234

Kirk, L and Burton, M (1977) Meaning and context: A study of conceptual shifts in meaning of Maasai personality descriptors. American Ethnologist, 4 (4), 734-761

Kirkland, J., Bimler, D. and Leonti, M. (1992) Cryograms and clinical applications. Paper presented at First International Conference, "Clinical implications of infant cry research".

Kirkland, J., Bimler, D. and Leonti, M. (1993) A method for surveying and mapping perceived events. Unpublished manuscript, Massey University

Klingberg, F. L. (1941) Studies in measurement of the relations among sovereign states. Psychometrika, 6 (6), 335-352

Kosslyn S. M., Pick, H. L. Jnr. and Fariello, G. R. (1974) Cognitive maps in children and men. Child Development, 45, 707-716

Krantz, D. H. (1967) Small-step and large-step color differences for monochromatic stimuli of constant brightness. Journal of the Optical Society of America, 57 (11), 1304-1316

Kraus, V., Schild E. O. and Hodge, R. W. (1978) Occupational prestige in the collective conscience. Social Forces, 56, 900-918

Krumshansl, C. L., Bharucha, J. J. and Kessler, E. J. (1982) Perceived harmonic structure of chords in three related musical keys. Journal of Experimental Psychology: Human Perception and Performance, 8, 24-36

Krumshansl, C. L. and Kessler, E. J. (1981) Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. Psychological Review, 89, 334-368

Kruskal, J. B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-27

Landis, D., Silver, C. A., Jones, J. M. and Messick, S. (1967) Level of proficiency and multi-dimensional viewpoints about problem similarity. Journal of Applied Psychology, 51 (3), 216-222

Lawless, H. T. (1989) Exploration of fragrance categories and ambiguous odors using multidimensional scaling and cluster analysis. Chemical Senses, 14 (3), 349-360

Levelt, W. J. M., van der Geer J. P. and Plomp, R. (1966) Triadic comparisons of muscial intervals. British Journal of Mathematical and Statistical Psychology, 19 (2), 163-179

Lingoes, J. C. and Borg, I. (1978) A direct approach to individual differences scaling using increasingly complex transformations. Psychometrika, 43 (4), 491-519

Livingstone, M. S. (1988) Art, illusion and the visual system. Scientific American, Jan. 1988, 68-75

Lopes, L. L. and Oden, G. C. (1980) Comparison of two models of similarity judgment. Acta Psychologica, 46, 205-234

Lutz, C. (1982) The domain of emotion words on Ifaluk. American Ethnologist, 9, 113-128

Lyons, P. J., Kirkland, J., Castle, O. and Lawoko, C. (1991) A computer-based method for triads: Assisting human judges choose between discrete items. Unpublished manuscript, Massey University

McAdams, S. and Cunible, J.-C. (1992) Perception of Timbral Analogies. Philosophical Transactions of the Royal Society of London B, 336, 383-389

McQuitty, L, L, and Clark, J. A. (1968) Multiple clusters, types and dimensions for iterative, intercolumnar correlation analysis. Multivariate Behavioral Research, 3, 465-477

MacRae, A. W., Howgate, P., and Geelhoed, E. (1990) Assessing the similarity of odours by sorting and by triadic comparison. Chemical Senses, 15 (6), 691-699

MacRae, A. W., Rawcliffe, T., Howgate, P., and Geelhoed, E. (1992) Patterns of odour similarity among carbonyls and their mixtures. Chemical Senses, 17 (2) 119-125

Mathews, M. W. and Pierce, J. R. (1980) Harmony and nonharmonic partials. Journal of the Acoustical Society of America, 68, 1252-1257

Melzack, R. and Torgerson, W. S. (1970) On the language of pain. Anesthesiology, 34, 50-59

Messick, S. J. (1956) An empirical evaluation of multidimensional successive intervals. Psychometrika, 21 (4), 367-375

Meulman, J. J. and Verboon, P. (1993) Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. Psychometrika, 58 (1), 7-35

Michon, J. A. (1972) Multidimensional and hierarchical analysis of progress in learning. In: Cognition in Learning and Memory (ed. L. W. Gregg). London: John Wiley

Miller, G. A. (1969) A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 6, 169-191

Miller, G. A. (1971) Empirical methods in the study of semantics. In: Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology (eds. Steinberg & Jakobovits), Cambridge, Cambridge University Press.

Miller, G. A. and Nicely. P. E. (1955) An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27 (2), 338-352

Miller, K. and Gelman, R. (1983) The child's representation of number: A multidimensional scaling analysis. Child Development, 54, 1470-1479

Miller, M. L. (1974) A comparison of judged-similarity, trait inference and trait rating tasks with regard to the multidimensional structure of personality traits. Unpublished Ph.D. thesis, University of California, Irvine. University Microfilms order number 75-11,034

Moore, R., Miller M. L, Weinstein, P., Dworkin, S. F. and Liou, H.-H. (1986) Cultural perceptions of pain and pain coping among patients and dentists. Community Dental and Oral Epidemiology, 14 (6), 327-333

Moore, R. and Dworkin, S. F. (1988) Ethnographic methodologic assessment of pain perceptions by verbal description. Pain, 34, 195-204

Morley, S. (1989) The dimensionality of verbal descriptors in Tursky's pain perception profile. Pain, 37, 41-49

Mosteller, F. (1958) The Mystery of the Missing Corpus. Psychometrika, 23, 279-289

Neidell, L. A. (1972) Procedures for obtaining similarities data. Journal of Marketing Research, 9, 335-337

Nerlove, S. B. and Burton, M. L. (1972) A further examination of cognitive aspects of English kinship terms. American Anthropologist, 74 (5), 1249-1253

Nordlie, P. G. (1958) A longitudinal study of interpersonal attraction in a natural group setting. Unpublished Ph. D. thesis. University Microfilms order number 58-7775

Nummenmaa, T. (1990) Sender repertoires of pure and blended facial expressions of emotion. Scandinavian Journal of Psychology, 31, 161-180

Osgood, C. E. (1966) Dimensionality of the semantic space for communication via facial expressions. Scandinavian Journal of Psychology, 7, 1-30

Paddick, R. G. (1978) Multidimensional scaling of perceived odour using structurally related odorants. Unpublished Ph.D. thesis, University of Canterbury.

Plomp, R. (1970) Timbre as a multidimensional attribute of complex tones. In: Frequency Analysis and Periodicity Perception in Hearing (eds. Plomp & Smoorenburg), Leiden: Sijthoff, pp. 397-414

Plomp, R. and Levelt, W. J. M. (1965) Tonal consonance and critical bandwidth. Journal of the Acoustical Society of America, 38, 548-560

Plomp, R. and Steeneken, H. J. M. (1969) Effect of phase on the timbre of complex tones. Journal of the Acoustical Society of America, 46 (2), 409-421

Plomp, R. and Steeneken, H. J. M. (1973) Place dependence of timbre in reverberant sound fields. Acustica, 28, 50-59

Plomp, R., Wagenaar, W. A. and Mimpen, A. M. (1973) Musical interval recognition with simultaneous tones. Acustica, 29, 101-109

Pollard-Gott, L. Emergence of thematic concepts in repeated listening to music. Cognitive Psychology, 15, 66-94

Pollack, I. (1978) Decoupling of auditory pitch and stimulus frequency: The Shepard demonstration, revisited. Journal of the Acoustical Society of America, 63, 202-206

Pols, L. C. W., van der Kamp, L. J. Th. and Plomp, R. (1969) Perceptual and physical space of vowel sounds. Journal of the Acoustical Society of America, 46, 458-467

Pruzanksy, S. (1970) Judgments of similarities among initial consonants using an auditory sorting apparatus. Paper J2, 80th meeting, Acoustical Society of America

Pruzansky, S., Tversky, A. and Carroll, J. D. (1982) Spatial versus non-spatial representations of proximity data. Psychometrika, 47 (1), 3-24

Ramsay, J. O. (1977) Maximum likelihood estimation in multidimensional scaling. Psychometrika, 42, 241-266

Ramsay, J. O. (1978) Confidence regions for multidimensional scaling analysis. Psychometrika, 43 (2), 145-160

Ramsay, J. O. (1980) The joint analysis of direct ratings, pairwise preferences, and dissimilarities. Psychometrika, 45 (2), 149-165

Rao, V. R. and Katz, R. (1971) Alternative multidimensional scaling methods for large stimulus sets. Journal of Marketing Research, 8, 488-494

Rapoport, A (1967) A comparison of two tree-construction methods for obtaining proximity measures among words. Journal of Verbal Learning and Verbal Behavior, 6, 884-890

Rapoport, A. and Fillenbaum, S. (1972) An experimental study of semantic structures. In: Multidimensional Scaling: Theory and applications in the Behavioral Sciences, Volume II (eds. R. N. Shepard, A. K., Romney, S. B. Nerlove). NY: Academic Press

Reading, A. E., Everitt, B. S. and Sledmere, C. M. (1982) The McGill Pain Questionaire: A replication of its construction. British Journal of Clinical Psychology, 21, 339-349

Richards, J. T. and Hanson, V. I. (1985) Visual and production similarity of the handshapes of the American manual alphabet. Perception and Psychophysics, 38 (4), 311-319

Richardson, M. W. (1938) Summary of address to American Psychological Association. Psychological Bulletin, 35, p.659

Rips, L. J., Shobe, E. J. and Smith, E. E. (1973) Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 12, 1-20

Robinson, J. P. and Hefner, R. (1967) Multidimensional differences in public and academic perception of nations. Journal of Personality and Social Psychology, 7, 251-259

Romney, A. K. and D'Andrade, R. G. (1964) Cognitive aspects of English kin terms. American Anthropologist, 66 (3 part 2) 146-170 (special publication)

Rosenberg, S. (1982) The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In: Multivariate Applications in the Social Sciences (eds. N. Hirschberg and L. G. Humphreys). NJ: Lawrence Erlbaum

Rosenberg, S. and Gara, M. A. (1983) Contemporary perspectives and future directions of personality and social psychology. Journal of Personality and Social Psychology, 45 (1), 57-73

Rosenberg, S. and Kim, M. P. (1975) The method of sorting as a data-gathering procedure in multivariate research. Multivariate Behavioral Research, 10, 489-502

Rosenberg, S., Nelson, C. and Vivekananthan, P. S. (1868) A multidimensional approach to the structure of personality impressions. Journal of Personality and Social Psychology, 9 (4), 283-294

Roskam, E. E. (1968) Metric analysis of ordinal data in psychology. Vam-Voorschoten.

Roskam, E. E. (1970) The method of triads for nonmetric multidimensional scaling. Nederlands Tijdschritt Voor de Psychologie en haar grensgebieden, 25, 404-417

Ross, J. (1966) A remark on Tucker and Messick's "Points of View" analysis. Psychometrika, 31 (1), 27-31

Rumelhart, D. E. and Abrahamson, A. A. (1973) A model for analogical reasoning. Cognitive Psychology, 5, 1-28

Russell, J. A. (1980) A circumplex model of affect. Journal of Personality and Social Psychology, 39, 1161-1178

Russell, J. A. (1983) Pancultural aspects of the human conceptual organization of emotions. Journal of Personality and Social Psychology, 45 (6), 1281-1282

Russell, J. A. and Bullock, M. (1985) Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. Journal of Personality and Social Psychology, 48, 1290-1298

Russell, J. A. and Bullock, M. (1986) On the dimensions preschoolers use to interpret facial expressions of emotion. Developmental Psychology, 22 (1), 97-102

Russell, J. A. and Fehr, B. (1987) Relativity in the perception of emotion in facial expressions. Journal of Experimental Psychology: General, 116 (3), 223-237

Russell, J. A., Lewicka, M. and Niit, M. (1989) A cross-cultural study of a circumplex model of affect. Journal of Personality and Social Psychology, 57 (5), 848-856

Russell, J. A. and Pratt, G. (1980) A description of the affective quality attributed to enviroments. Journal of Personality and Social Psychology, 38 (2), 311-322

Sampson, S. F. (1968) A novitiate in a period of change: An experimental and case study of social relationships. Unpublished Ph.D. thesis, Cornell University. University Microfilms 69-5775

Schiffman, S. S., Reynolds, M. L. and Young, F. W. (1981) Introduction to Multidimensional Scaling: Theory, Methods and Applications. NY: Academic Press

Schmidt, C. F. (1972) Multidimensional analysis of the printed media's explanations of the riots of the summer of 1967. Journal of Personality and Social Psychology, 24 (1), 59-67

Schneider, B. A. and Bissett, J. (1981) The dimensions of tonal experience: A nonmetric multidimensional scaling approach. Perception and Psychophysics, 30 (1), 39-48

Schneider, B. A., Scott, P. and Stein, D. (1974) The measurement of loudness using direct comparisons of sensory intervals. Journal of Mathematical Psychology, 11, 259-273

Seitz, V. R. (1971) Multidimensional scaling of dimensional preferences: A methodological study. Child Development, 42, 1701-1720

Shepard, R. N. (1962) The analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika, 27 (3), 219-246

Shepard, R. N. (1964) Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1, 54-87

Shepard, R. N. (1964) Circularity in judgments of relative pitch. Journal of the Acoustical Society of America, 36 (12), 2346-2353

Shepard, R. N. (1974) Representation of structure in similarity data: Problems and prospects. Psychometrika, 39 (4), 373-421

Shepard, R. N. (1978) The circumplex and related topological manifolds in the study of perception. In: Theory construction and data analysis in the behavioral sciences (ed. S. Shye). SF: Jossey-Bass

Shepard, R. N. and Cermak, G. W. (1973) Perceptual-cognitive explorations of a toroidal set of free-form stimuli. Cognitive Psychology, 4, 351-377

Shepard, R. N. and Chipman, S. (1970) Second order isomorphism of internal representations: Shapes of states. Cognitive Psychology, 1, 1-17

Shepard, R. N. and Cooper, L. A. (1992) Representation of colors in the blind, colour-blind, and normally sighted. Psychological Science, 3 (2), 97-104

Shepard, R. N., Kilpatric, D. W. and Cunningham, J. P. (1975) The internal representation of numbers. Cognitive Psychology, 7, 82-138

Sherman, R. C. (1972) Individual differences in perceived trait relationships as a function of dimensional salience. Multivariate Behavioral Research, 7, 109-129

Shweder, R. A. (1972) Semantic structures and personality assessment. Unpublished Ph.D. thesis, Harvard University. University Microfilms order number 72-29,584

Silver, C. A., Landis, D. and Messick, S. (1966) Multidimensional analysis of visual form: An analysis of individual differences. American Journal of Psychology, 79, 62-72

Slaymaker, F. H. (1970) Chords from tones having stretched partials. Journal of the Acoustical Society of America, 47, 1569-1571

Spence, I. and Domoney, D. W. (1974) Single subject incomplete designs for nonmetric multidimensional scaling. Psychometrika, 39 (4), 469-490

Stalmeier, P. F. M. and de Weert, C. M. M. (1991a) Large colour differences measured by spontaneous gestalt formation. Color Research and Application, 13 (4), 209-218

Stalmeier, P. F. M. and de Weert, C. M. M. (1991b) Large color differences and selective attention. Journal of the Optical Society of America A, 8 (1), 237-247

Stalmeier, P. F. M. and de Weert, C. M. M. (1994) On the conditions affecting the contribution of color to brightness perception. Color Research and Application, 19 (3), 192-201

Steinberg, D. (1967) The word sort: An instrument for semantic analysis. Psychonomic Science, 12, 541-542

Steinheiser, F. H. Jnr. (1970) Individual preference scales within a multidimensional 'similarities' space. Journal of Experimental Psychology, 86, 325-327

Stephenson, W. (1953) The Study of Behavior: Q-Technique and its Methodology. Chicago: University of Chicago Press

Stoop, I. (1986) The contribution of merging: A comparison of two data collection methods in MDS. Abstracted in: Multidimensional Data Analysis, Leiden, DSWO Press

Stringer, P. (1967) Cluster analysis of non-verbal judgments of facial expressions. British Journal of Mathematical and Statistical Psychology, 20, 71-79

Struhsaker, T. T. (1967) Social structure among vervet monkeys (*cercopithecus æthiops*). Behaviour, 29, 83-121

Takane, Y. (1978) A Maximum Likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent – Theory. Japanese Psychological Research, 20 (1), 7-17

Takane, Y. (1978) A Maximum Likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent – Evaluations. Japanese Psychological Research, 20 (3), 105-114

Takane, Y. (1980) Analysis of categorizing behavior by a quantification method. Behaviormetrika, 8, 75-86

Takane, Y. (1981) MDSORT: A special-purpose MDS program for sorting data. Behavioral Research Methods and Instrumentation, 13 (5), p.698

Takane, Y. (1981) Multidimensional successive categories scaling: A Maximum Likelihood method. Psychometrika, 46, 9-28

Takane, Y. and Carroll, J. D. (1981) Nonmetric Maximum Likelihood multidimensional scaling from directional ranking of similarities. Psychometrika, 46, 389-405

Takane, Y., Young, F. W. and de Leeuw, J. Nonmetric individual differences multidmensional scaling: An Alternating Least Squares method with optimal scaling features. Psychometrika, 42 (1), 7-67

Tansley. B. W. and Boynton, R. M. (1976) A line, not a space, represents visual distinctness of borders formed by different colors. Science, 191, 954-957

Thompson, P. (1983) Some missing data patterns for multidimensional scaling. Applied Psychological Measurement, 7 (1), 44-55

Torgerson, W. S. (1952) Multidimensional scaling: I. Theory and Method. Psychometrika, 17, 401-419

Torgerson, W. S. (1958) Theory and Methods of Scaling. NY: John Wiley

Torgerson, W. S. and BenDebba, M. (1983) The structure of pain descriptors. In: Pain Measurement and Assessment (ed. Melzack) New York, Raven Press

Truex, G. F. (1973) An experimental study of the semantic structure of some Zapotec verbs. Unpublished Ph.D. thesis, University of California, Irvine. University Microfilms 73-31,441

Tucker, L. R. and Messick, S. (1963) An individual differences model for multidimensional scaling. Psychometrika, 28 (4), 333-367

Tuck, D. C., Wack, J. T. and Kerns, R. D. (1985) An empirical examination of the 'Pain-Behavior' construct. Journal of Behavioral Medicine, 8 (2), 119-130

Tyska, T. and Goszczynska, M. (1993) What verbal reports say about risk perception. Acta Psychologica, 83, 53-64

van der Kamp, L. J. Th. and Pols, L. C. W. (1971) Perceptual analyses from confusions between vowels. Acta Psychologica, 35, 64-77

van der Kloot, W. A. and van Herk, H. (1991) Multidimensional scaling of sorting data: A comparison of three procedures. Multivariate Behavioral Research, 26 (4), 563-381

Verkes, R. J., van der Kloot, W. A. and Van der Meij, J. (1989) The perceived structure of 176 pain descriptive words. Pain, 38, 219-229

Vlaeyen, J. W. S. , van Eek, H., Groenman, N. H. and Schuerman, J. A. (1987) Dimensions and components of observed chronic pain behaviour. Pain, 31, 65-85

Vlek, C. and Stallen, P. J. (1981) Judging risks and benefits in the small and in the large. Organizational Behavior and Human Performance, 28, 235-271

Wack, J. T. and Turk, D. C. (1984) Latent structure of strategies used to cope with nociceptive stimulation. Health Psychology, 3, 27-43

Ward, L. M. (1977) Multidimensional scaling of the molar physical environment. Multivariate Behavioral Research, 12, 23-42

Weinberg, S. L. and Menil, V. C. (1993) The recovery of structure in linear and ordinal data: INDSCAL versus ALSCAL. Multivariate Behavioral Research, 28 (2), 215-233

Wessel, D. L. (1979) Timbre space as a musical control structure. Computer Music Journal, 3 (2), 45-52

Wexler, K. N. and Romney, A. K. (1972) Individual variations in cognitive structures. In: Multidimensional Scaling: Theory and applications in the Behavioral Sciences (eds. R. N. Shepard, A. K. Romney, S. B. Nerlove)

White, G. (1978) Ambiguity and ambivalence in A'ara personality descriptors. American Ethnologist, 5, 334-360

Winsberg, S. and Carroll, J. D. (1989) A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. Psychometrika, 54 (2), 217-229

Winsberg, S. and de Soete, G. (1993) A latent class approach to fitting the weighted Euclidean model, CLASCAL. Psychometrika, 58 (2), 315-330

Winton, W., Ough, C. S. and Singleton, V. L. (1975) Relative distinctiveness of varietal wines estimated by the ability of trained panelists to name the grape variety correctly. American Journal of Enology and Viticulture, 26 (1), 5-11

Wish, M. and Carroll, J. D. (1974) Applications of individual differences scaling to studies of human perception and judgment. In: Handbook of Perception, Volume II (eds E. C. Carterette and M. P. Friedman). NY: Academic Press

Wright, A. W. and Cumming, W. W. (1971) Color-naming functions for the pigeon. Journal of the Experimental Analysis of Behavior, 15, 7-17

Wright, H. (1965) Precision of color differences derived from a multidimensional scaling experiment. Journal of the Optical Society of America, 55 (12), 1650-1655

Yoshida, M. (1975) Psychometric classification of odors. Chemical Senses and Flavor, 1, 443-464

Young, F. W. (1975) Scaling replicated conditional rank-order data. Sociological Methodology 6 (ed. D. R. Heise), 129-170

Young, F. W. and Cliff, N. (1972) Interactive scaling with individual subjects. Psychometrika, 37 (4), 385-415

Young, F. W., Null, C. H. and Sarle, W. (1978) Interactive similarity ordering. Behavioral Research Methods and Instrumentation, 10 (2), 273-280

Young, G. and Householder, A. S. (1938) Discussion of a set of points in terms of their mutual distance. Psychometrika, 3, 19-22