An evaluation of the use of multidimensional scaling for understanding brain connectivity

GEOFFREY J. GOODHILL*, MARTIN W. SIMMEN AND DAVID J. WILLSHAW

University of Edinburgh, Centre for Cognitive Science, 2 Buccleuch Place, Edinburgh EH8 9LW, U.K.

SUMMARY

A large amount of data is now available about the pattern of connections between brain regions. Computational methods are increasingly relevant for uncovering structure in such datasets. There has been recent interest in the use of non-metric multidimensional scaling (NMDS) for such analysis. NMDS produces a spatial representation of the 'dissimilarities' between a number of entities. Normally, it is applied to data matrices containing a large number of levels of dissimilarity, whereas for brain connectivity data there is a very small number. We address the suitability of NMDS for this case. Systematic numerical studies are presented to evaluate the ability of this method to reconstruct known geometrical configurations from dissimilarity data possessing few levels. In this case there is a strong bias for NMDS to produce annular configurations, whether or not such structure exists in the original data. For the case of a connectivity dataset derived from the primate cortical visual system, we demonstrate that great caution is needed in interpreting the resulting configuration. Application of an independent method that we developed also strongly suggests that the visual system NMDS configuration is affected by an annular bias. We question the strength of support that an NMDS analysis of the visual system data provides for the two streams view of visual processing.

1. INTRODUCTION

The pattern of interconnectivity between different regions of the vertebrate brain provides clues and constraints to hypotheses about brain function. Considerable time and effort has been put into experiments to determine the afferent and efferent patterns for different regions. From such experiments datasets of great size and complexity are being generated, and computational methods for uncovering and visualizing the structure within these datasets are becoming increasingly relevant. In this paper we discuss the appropriateness of one general method of data analysis/visualization, that of multidimensional scaling (MDS) (Torgerson 1952; Shepard 1962*a*,*b*; Kruskal 1964*a*,*b*), for understanding brain connectivity. We focus on the dataset of connections between areas in the primate visual cortex that have been identified by anatomical and physiological criteria (Felleman & Van Essen 1991). There have been many attempts to derive connectivity maps for different parts of the brain, which have led to wiring diagrams of ever increasing complexity. Young (1992) was the first to apply MDS to this problem for the primate visual cortex.

In its simplest form, MDS takes as input a symmetric $N \times N$ matrix which describes the 'dissimilarities' between a set of N entities. We discuss the case where the entities are brain regions and the dissimilarities are information regarding whether or not these regions are connected. The output is a plot of N points, one for each entity, in a low-dimensional space (usually two- or three-dimensional for ease of interpretation). The points are positioned so that the distances between them reflect as closely as possible (in a particular sense) the dissimilarities between the entities given in the original matrix. In metric MDS (Torgerson 1952), the absolute values of the dissimilarities are taken to be meaningful, whereas in the non-metric version (NMDS) (Shepard 1962*a*,*b*; Kruskal 1964*a*,*b*) only the *ordering* of dissimilarities is considered. Here the aim is to find a configuration of points in which the rank order of distances between points reflects as closely as possible the rank order of the corresponding dissimilarities in the original matrix. There are many different extensions of MDS, for instance to non-Euclidean distance metrics and to asymmetric matrices, which will not concern us here. For reviews see Shepard (1980), Coxon (1982), Young (1984, 1987), Torgerson (1986), Gower (1987), Young & Harris (1990).

In this paper we focus on NMDS. This was developed in the 1960s for problems in psychology, where the aim was to learn for instance about a subject's internal representation of various stimuli (e.g. colours) by studying the degree to which pairs of stimuli are judged to be similar (see, for example, Shepard 1980). Since then both NMDS, and to a lesser extent metric MDS, have found widespread application in the biological sciences.

* Present address: The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, U.S.A.

Some illustrative examples are: (i) in taxonomy, to produce graphical representations of the relationships between species (Rohlf 1970; Jensen & Barbour 1981); (ii) in ecology, to visualize relations in large species-by-site datasets (Legendre & Legendre 1987; Digby & Kempton 1987); (iii) in molecular biology, for sequence analysis (Higgins 1992; Hess *et al.* 1994); and (iv) in neurophysiology, for understanding population responses of neurons in the monkey visual cortex (Hasselmo *et al.* 1989; Young & Yamane 1992). In addition, in the field of neural networks a number of algorithms have been developed for mapping high-dimensional data into a low-dimensional space (e.g. Kohonen 1982; Durbin & Willshaw 1987; Durbin & Mitchison 1990). Relations between these algorithms and MDS are beginning to be explored (Lowe 1993). However, despite a great deal of discussion of the theoretical underpinnings of MDS in the 1960s and 1970s (see the reviews cited above), many questions regarding the best way to apply the method, and its validity for particular types of data, remain unresolved. Many people have argued (see, for example, the discussion following Ramsay (1982)), that the proper place for MDS is as a tool for data exploration rather than analysis.

Perhaps the most important difference between the use of NMDS for brain connectivity data compared to more conventional applications is that, from the currently available data, it is difficult to make systematic comparisons of the relative strengths of different connections. The most robust way to use this data is to regard it as providing *binary* information about dissimilarities: a connection exists in either direction (similar) or it does not (dissimilar). However, as NMDS tries to reflect the *ordering* in the data, it is reasonable to question whether it can produce meaningful results when applied to matrices containing only a small number of distinct levels of dissimilarity: just two in the binary case. Very little analysis or empirical study has been undertaken regarding the validity of NMDS under these conditions. Here we address this issue in a number of cases. We show that a systematic bias occurs which introduces artefactual structure into the results, and that this is highly relevant to the conclusions which it is safe to draw from the application of NMDS to connectivity data. This conclusion is of biological significance since NMDS is starting to be applied to a variety of connectivity datasets, and inferences of a biological nature are being drawn from the resulting configurations (Young 1992, 1993; Scannell & Young 1993; O'Mara *et al.* 1994).

The structure of the paper is as follows. We first introduce the visual system dataset, and the so-called 'two streams' hypothesis of visual processing. NMDS is then applied to a binary similarity matrix derived from this data to give a two-dimensional configuration of areas, and how this configuration might seem to provide strong support for the two streams view is discussed. We then go on to an investigation of the appropriateness of NMDS for binary data more generally. We start by introducing the details of the NMDS method which are necessary for understanding the results that follow. We then present an empirical study in three parts. First, we show illustrative results for two example configurations. Second, a Monte Carlo study is presented of the performance of NMDS in reconstructing a large number of artificially constructed configurations from binary dissimilarity data in terms of various measures of both real and apparent fit. Third, we study the overall geometry of the derived configuration, in order to examine the influence of artefactual structure. In the light of these results, we return to the configuration derived for the visual system data. Based on our NMDS analysis, and also on an independent analysis, we argue that this configuration contains a significant proportion of artefactual structure, which greatly affects its interpretation with respect to the two streams hypothesis. Finally we address some counter-arguments to our case that have recently been put forward, and we consider some alternative ways in which NMDS could be applied to this problem. A preliminary summary of some of this work has already appeared in Simmen *et al.* (1994).

2. APPLICATION OF NMDS TO AREAS IN THE PRIMATE CORTICAL VISUAL SYSTEM

The primate cortical visual system can be divided into functionally distinct areas using various schemes. Felleman & Van Essen (1991) distinguish 32 areas, each of which projects to many other areas, to give an extremely complex wiring diagram (figure 4 in Felleman & Van Essen 1991). Perhaps the most important organizational hypothesis that has been put forward regarding the functional relations between cortical areas and the way in which information is transformed through this network of areas is that, starting from the primary visual cortex (V1), information flows out along two segregated streams, each of which is hierarchically organized. This 'two streams' hypothesis was first proposed by Ungerleider & Mishkin 1982 (see also Mishkin *et al.* (1983)) on the basis of anatomical, physiological and lesion data, and quickly became the dominant view of the gross organization of cortical visual processing (see, for example, Maunsell & Newsome 1987; Livingstone & Hubel 1988; Desimone & Ungerleider 1989; Baizer *et al.* 1991). More recently, the more extreme versions of this view have come increasingly under attack, as evidence accumulates for the intermixing of information between supposedly segregated pathways (e.g. DeYoe & Van Essen 1988; Zeki & Shipp 1988; Goodale & Milner 1992; Merigan & Maunsell 1993). It is of interest therefore to see whether the application of an objective method for analysing the connectivity data such as NMDS can shed light on these issues.

Our source data was the connectivity matrix forming table 3 of Felleman & Van Essen (1991), henceforth referred to as the FVE matrix. This contains information about both the afferent and efferent connection patterns for each area. Of $31 \times 32 = 992$ possible connections, 675 are indicated as having been established uncontroversially as present or absent. The pattern of connections is almost symmetric, as in all but five of the cases that have been investigated in both directions, the connections are reciprocal. From the FVE matrix we

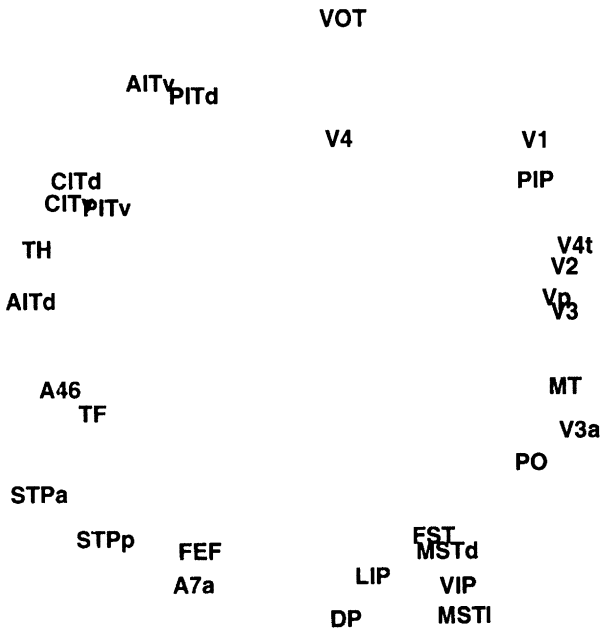

Figure 1. The configuration derived by ALSCAL from a binary similarity version of the FVE matrix. This has an RSQ of 0.47 and a SSTRESS value of 0.41 (see text for explanation of terms).

derived a symmetric binary similarity matrix for 30 of the areas, in which an entry is 1 if a connection exists in either or both directions. Following Young (1992) we omitted areas MIP and MDP because of the uncertainty about the connections they make with other areas of the visual cortex (Felleman & Van Essen 1991). We submitted this matrix to the standard NMDS program ALSCAL at the ordinal level of measurement and adopting the tied approach to ties (these terms will be explained in the next section) to derive the two-dimensional configuration shown in figure 1. This configuration has a strongly *annular* form. It suggests that, starting from V1, visual information flows out in two highly distinct and hierarchically organized streams as defined by the division and the ordering of the areas in the two arms of the annulus, and then reconverges in areas 46 and TF. The identity and ordering of areas in the two arms correspond roughly with that hypothesized in the two streams view for the temporal and parietal pathways. In earlier work, Young (1992) performed a closely related NMDS analysis and derived a very similar annular configuration. The main difference between Young's analysis and ours is that he derived a *ternary* similarity matrix from the FVE matrix, where each similarity is assigned value 2 if the corresponding areas are reciprocally connected, 1 if a connection in one direction exists (the pathway in the other direction being absent or untested), and 0 if both pathways are either absent or have not yet been tested for. We argue that a binary approach is more suitable for two reasons. Firstly, it is somewhat arbitrary to assert that areas connected in only one direction are 'less similar' than those connected in both directions. Secondly, the strong reciprocity constraint existing for connections that have been investigated in both directions suggests that it is more appropriate to assign the same value to an untested direction as to the tested direction, rather than assuming that all untested connections are absent. If this were done virtually no similarities of value 1 would be left in Young's similarity matrix, rendering it almost binary. Later we show that the bias towards annularity in NMDS configurations derived from binary data is in any case shared by NMDS configurations derived from ternary data. The RSS (see §3*b* (iii)) between our configuration and Young's is 0.045, indicating that they are in fact very similar. Young concluded that his configuration provides strong objective support for the two-streams view.†

However, this interpretation rests to a large extent on the strongly annular nature of the configuration. In the rest of this paper we present results to show that NMDS has an intrinsic bias to produce annular configurations given binary data, whether or not such structure is actually present in the data. To make our case it is necessary to go into some detail about the mechanics of NMDS in the next section. Readers wishing to study this only briefly should focus on §§3*c* and 3*d*, which illustrate the tendency towards annularity, before proceeding to §4 where figure 1 is re-evaluated in the light of these results.

3. MDS AND ITS APPLICATION TO BINARY DISSIMILARITY DATA

(*a*) *Metric* MDS

The seminal algorithm for forming a representation of the dissimilarities between a set of entities by distances within a geometric configuration of points, referred to in the literature as classical MDS, metric MDS, metric scaling, or principal coordinates analysis, was proposed by Torgerson (1952). This method is used to construct a geometric configuration, in a space of any chosen dimensionality, in which the interpoint distances approximate the dissimilarities. Metric scaling is related to principal components analysis, one difference being that the latter operates on an $N \times M$ multivariate data matrix whereas metric scaling operates on an $N \times N$ matrix of dissimilarities. When a dissimilarity matrix is constructed from a multivariate matrix by defining the dissimilarity between two entities to be the Euclidean distance between their respective locations in M-dimensional attribute space, both methods will produce the same configuration of points (Gower 1966).

If the dissimilarities are Euclidean distances, the dissimilarity matrix can be represented exactly by a configuration in a Euclidean space of dimension N-1. In practice N is large and the aim is to find the best approximation in a space of many fewer than N-1 dimensions. This can be achieved by calculating the eigenvalues and eigenvectors of a certain matrix derived from the dissimilarity matrix (an analytic one-step procedure). See Gower (1966) for details of the method. The best configuration in d-dimensional space

† We note that: (i) there are six discrepancies between Young's ternary matrix and the FVE matrix, namely the entries for PITd-MSTd, PITd-FST, PITv-FST, PITv-MSTd, FEF-VIP and A46-MT; and (ii) the published picture of the configuration obtained by Young (Young 1992, figure 1) is stretched horizontally by a factor of approximately 1.3; the proper coordinates were kindly supplied to us by Dr Young.

can then be computed from the leading d eigenvectors and eigenvalues, with the ratio of the sum of the first d eigenvalues to the sum of all the eigenvalues reflecting the goodness of fit. However, often the dissimilarities are non-Euclidean, leading to some negative eigenvalues. Provided none of these negative eigenvalues is large in magnitude, a real solution of low dimensionality may often still be obtainable. See Gower (1966) or Krzanowski (1988, pp. 104–113) for discussion of this issue.

Metric MDS assumes that dissimilarity data is at either the ratio or interval level of measurement, rather than the weaker ordinal level.† We next describe an algorithm that is free from this limitation.

(*b*) *Non-metric* MDS

Various algorithms have been proposed to find the best geometric representation of dissimilarity data specified at the ordinal level (Shepard 1962*a*,*b*; Kruskal 1964*a*,*b*). We refer to the basic non-metric method as NMDS.

For any set of dissimilarities in a symmetric $N \times N$ matrix it is always possible to find a configuration in N-1 dimensions where the ordering of distances between points reflects the ordering of dissimilarities (Bennett & Hays 1960; Shepard 1962*a*). Again, the practical problem is to calculate the best approximation in a much lower dimensional space. The NMDS algorithms work by doing gradient descent in an objective function that specifies to what degree the rank ordering of distances in the configuration departs from the rank ordering of dissimilarities in the matrix. Our summary of the method follows Coxon (1982) and Krzanowski (1988). An initial configuration of points is chosen in the space of required dimensionality (often by applying metric scaling). The ordering of distances is then compared with the ordering of dissimilarities, and discrepancies identified. For each distance, a target value is defined such that, if all distances reached their targets, the two orderings would match. These target values are termed *disparities*. Many disparities may be equal: only weak monotonicity of disparities with dissimilarities is enforced.

An *objective function* of the discrepancies between distances and disparities is then minimized either directly, or by gradient descent. New disparities are then calculated, and the procedure iterates until there is negligible improvement in the objective function. The calculation of disparities can be expressed as the minimization of the same objective function when the variables are the disparities, given the constraint that they vary monotonically with the corresponding dissimilarities. There is no guarantee that a global rather than local minimum will be reached. The axes and scale of the final plot computed by NMDS are arbitrary: usually the plot is scaled by placing the centre of mass of the points at the origin and constraining the coordinate variance.

Several versions of NMDS exist, each using a different objective function or minimization procedure. Each algorithmic development has become bound up with a particular computer package (Gower 1982; Torgerson 1986), and so it is only possible to reproduce NMDS results if the same version of a computer package is used. Although one can make general claims about the behaviour of NMDS on a particular problem, behaviour at a detailed level is specific to the program used. Our results were obtained with the NMDS option of the widely used ALSCAL algorithm (Takane *et al.* 1977), as implemented in Version 4 of the SPSS statistical software package (Young & Harris 1990).

(i) *Objective functions*

The objective function for NMDS proposed by Kruskal (1964*a*,*b*) is termed STRESS. The square of STRESS is the sum over all pairs of points of the squared difference between distance and disparity, divided by the sum of squared distances. The function minimized by ALSCAL is called SSTRESS. The square of SSTRESS is the sum over squared differences between squared distances and squared disparities, divided by the sum of quartic disparities. This function was adopted for computational convenience, since it allows minimization to be performed by an efficient 'alternating least squares' algorithm rather than steepest descent. When interpreting ALSCAL output, it should be borne in mind that SSTRESS emphasizes the fitting of large disparities over small ones (see, for example, Greenacre & Underhill 1982). Empirically, the STRESS values of scaling solutions have been found to be approximately linearly related to the SSTRESS values, with proportionality constant ≈ 0.75 (Young & Null 1978); a similar relation was found to hold for the solutions generated in the current study (data not shown). Another quantity sometimes quoted as a performance measure for NMDS, particularly for ALSCAL, is RSQ, the squared correlation between disparities and distances. For a perfect solution RSQ is 1.0. Note that RSQ is not optimized explicitly. We shall often call SSTRESS and RSQ 'apparent fit' measures.

Considerable effort has gone into quantitatively interpreting STRESS (and SSTRESS) values when, as is usually the case, the underlying configuration (if one exists) is unknown. One approach is to compare the observed STRESS value with those obtained by scaling random dissimilarity matrices possessing no structure: if the former lies within the distribution of the latter then it is safe to conclude that the original dissimilarities were effectively noise and thus that the derived configuration be disregarded (Klahr 1969; Stenson & Knoll 1969). However, this approach is of limited utility, since even if the hypothesis of randomness can be rejected '... the data may be so errorful that the resulting solution is relatively useless' (Spence & Ogilvie 1973, p. 516). An alternative approach, utilized in this paper, is to scale datasets generated from

† Conventionally, four levels of measurement are defined (Stevens 1951). *Ratio*: absolute values are meaningful; both the intervals and zero point of the scale are relevant (example: mass). *Interval*: intervals in the data are meaningful, but not the zero point (example: Celsius temperature scale). *Ordinal*: only the ordering of the data is meaningful, not the values (example: a subject's rating of taste stimuli in a psychology experiment). *Nominal*: the data describes only distinct unordered categories.

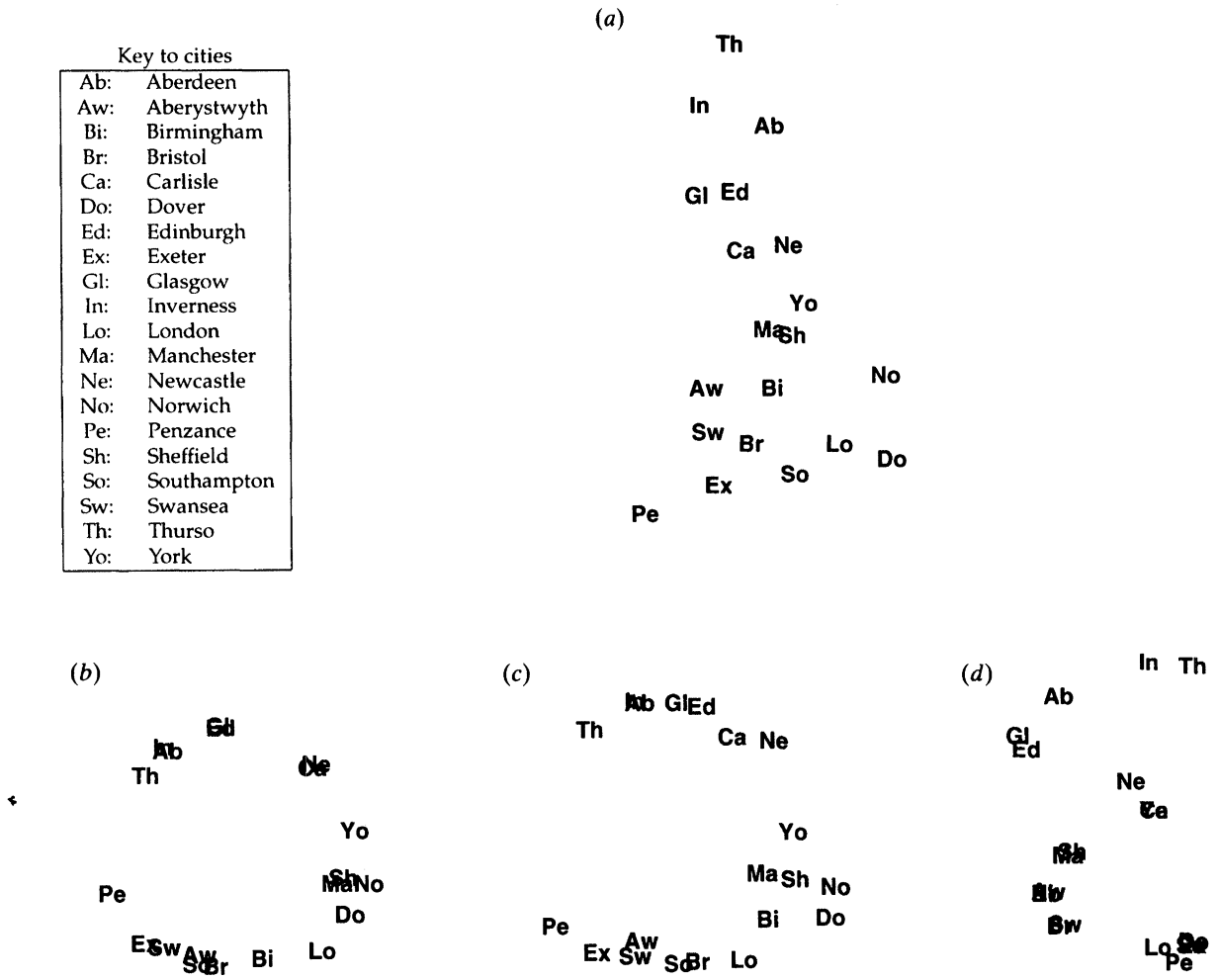

Figure 2. Configurations derived from the city matrix using NMDS. (*a*) Using full dissimilarities. SSTRESS = 0.0011, RSQ = 1.000, RSS = 0.00012. (*b–d*) Using binarized matrix for $P_N = 30\%$, 49% and 79% respectively. SSTRESS, RSQ and RSS values are plotted in figure 3. Note that for large P_N some points become coincident as they now have identical rows in the dissimilarity matrix.

artificial geometric configurations. By incorporating an error model into the process of generating the dissimilarities, such Monte Carlo studies (e.g. Young 1970; Wagenaar & Padmos 1971; Sherman 1972) attempt to mimic real situations in which the experimental data reflects an underlying structure but is corrupted by noise. Comparison of the observed value of the objective function with the Monte Carlo values can suggest whether the solution represents valid structure.

(ii) *Tied data*

Two or more different entries in the dissimilarity matrix may have the same value. This causes ambiguity in defining an ordering of the data. Two approaches for dealing with tied data are usually considered (Kruskal 1964*a*): (i) untied approach: tied dissimilarity values may be broken in the configuration, i.e. there is no penalty if the corresponding distances are unequal; (ii) tied approach: tied values should be conserved as far as possible. These two approaches to ties are also referred to as primary and secondary. We prefer the untied/tied nomenclature since it is more explicit, and does not imply an ordering of the two methods. Most programs allow either approach. With a large number of distinct levels in the data, which approach is chosen has little effect on the final configuration. For a small number of levels the differences can be significant. There are no hard-and-fast rules for deciding which approach to ties is more applicable for a particular problem.

(iii) *Comparing configurations*

In comparing two configurations, only the internal relations between the points are relevant. We used Procrustes analysis (Schönemann & Carroll 1970; Gower 1971*b*; Sibson 1978). Here one configuration is transformed so that the residual sum of squared distances (RSS) between corresponding points is minimized. The transformations allowed are rigid translation, isotropic scaling, rigid rotation, and reflection. Given that both configurations are normalized, the value of RSS can range from 0.0, for a perfect fit, upwards to 1.0. The NAG (1991) subroutine G03BCF was used to perform the Procrustes analyses reported in this paper.†

† An alternative approach is to measure the correlation between the distances separating corresponding pairs of points in the two configurations (Shepard 1966).

Procrustes analysis is particularly useful in control studies when the true configuration underlying the dissimilarities is known; the success of the reconstruction by NMDS (or indeed any other method) can then be measured directly, rather than indirectly in terms of SSTRESS, STRESS or RSQ. In such studies it has been found that the apparent fit measures are not always good indicators of the real fit as given by measures of configuration similarity (e.g. Young 1970). Sometimes we will refer to 'the RSS value' of a NMDS configuration, meaning the value of RSS between that configuration and the true configuration. We will sometimes refer to RSS as a 'true fit' measure.

(*c*) *Reconstructing geometrical configurations from binary data*

A problem in assessing the quality of constructions obtained by NMDS is that usually either a true configuration is unknown or does not exist. This does not arise when test data is generated from geometrical configurations. We now investigate how accurately known configurations can be reconstructed by NMDS given only binary information about distances. Results are firstly shown for two particular configurations chosen to give a feeling for how well reconstruction can be achieved. These are: (i) a configuration of 20 cities in Great Britain, where inter-city dissimilarities were derived from Euclidean distances between cities on a map; and (ii) a regular square grid of 8×8 points, where inter-point dissimilarities are again Euclidean distances. We then present statistics summarizing a systematic study of an ensemble of randomly generated distributions of points in the plane. Here we concentrate on the tied approach: later we discuss the untied approach, and suggest that this offers little improvement.

In all cases binary dissimilarity matrices were generated by choosing a threshold distance and then setting all distances in the full dissimilarity matrix below threshold to zero ('near') and above to one ('far'). Several different binary matrices were generated from the same full dissimilarity matrix by choosing different threshold values to give different proportions of nears and fars. P_N is the percentage of nears in the resulting binary matrix (excluding the diagonal entries).

(i) *City and grid distributions*

We first applied ALSCAL to the full dissimilarity matrix. The true configuration was reconstructed in both cases: the resulting city configuration is shown in figure 2*a*. This demonstrates the basic power of the algorithm to produce accurate geometric reconstructions when given ordinal data with many distinct levels.

City configurations derived from binary matrices for a selection of values of P_N using the tied approach are shown in figure 2*b–d*. The overall structure of the true configuration is almost completely destroyed by the binarization for all values of P_N. This is reflected in figure 3, where SSTRESS, RSQ and RSS values are plotted as a function of P_N. According to all these criteria, performance is best for $P_N \approx 60\,\%$. Some configurations computed by NMDS for the grid distribution are shown in figure 4. As is apparent from both visual inspection and RSS values, the reconstruction is much better than for the city distribution, even though SSTRESS and RSQ values are comparable to those for the city distribution (figures 3 and 4).

The reason that in neither case does the RSQ approach 1 despite the success of the reconstruction is that there is an inherent upper bound on RSQ for scaling binary data using the tied approach (as first pointed out by Simmen 1994). The precise value of this RSQ 'ceiling' depends on the particular configuration obtained. To explain the origin of the ceiling, recall that RSQ is the squared correlation between distances and disparities (see §3*b*(i)) and for binary data there are just two disparities. In the optimal case there is no overlap between the sets of interpoint distances assigned to these two disparities. However, there is in general a spread within the two categories, which reduces to below 1 the correlation between distances and disparities. Only in the rare case of a configuration having just two distinct interpoint distances can the RSQ be 1. The RSQ ceiling for the city distribution is 0.71 (at $P_N = 67\,\%$), and for the grid distribution 0.70 (at $P_N = 55\,\%$). Thus in both cases the best RSQ achieved is close to the best possible. Similar arguments apply to dissimilarity data that consists of several discrete levels: the RSQ ceiling will approach 1 as the number of unique levels becomes large. Analogously, there is an SSTRESS 'floor', the existence of which can be seen in Young & Null (1978), although it was not remarked upon there.

Why is the reconstruction for grids generally better than for cities? The grid configuration has more points. It is well known that the true fit of NMDS solutions (sometimes called the degree of 'metric determinacy') tends to increase with the ratio of the number of elements in the dissimilarity matrix to the number of degrees of freedom in the derived configuration (Shepard 1966; Young 1970). This is borne out by the Monte Carlo study below. However, the more uniform distribution of points in the grid configuration may also aid accurate recovery.

(ii) *Random geometrical distributions*

We now present the results of a more systematic investigation involving many Monte Carlo experiments scaling binary dissimilarity data derived from a large number of artificially generated two-dimensional configurations. To be of help in interpreting the results of real NMDS analyses, ideally Monte Carlo studies should be conducted with configurations like those thought to underly the experimental data involved. Given that the underlying structure can usually only be guessed at, it is sensible to perform Monte Carlo studies on configurations of several different forms. Later we report Monte Carlo results obtained using configurations having annular form, but in this section we present Monte Carlo results obtained from what we call *disc* configurations. These were generated by randomly locating points within a circle of unit radius.

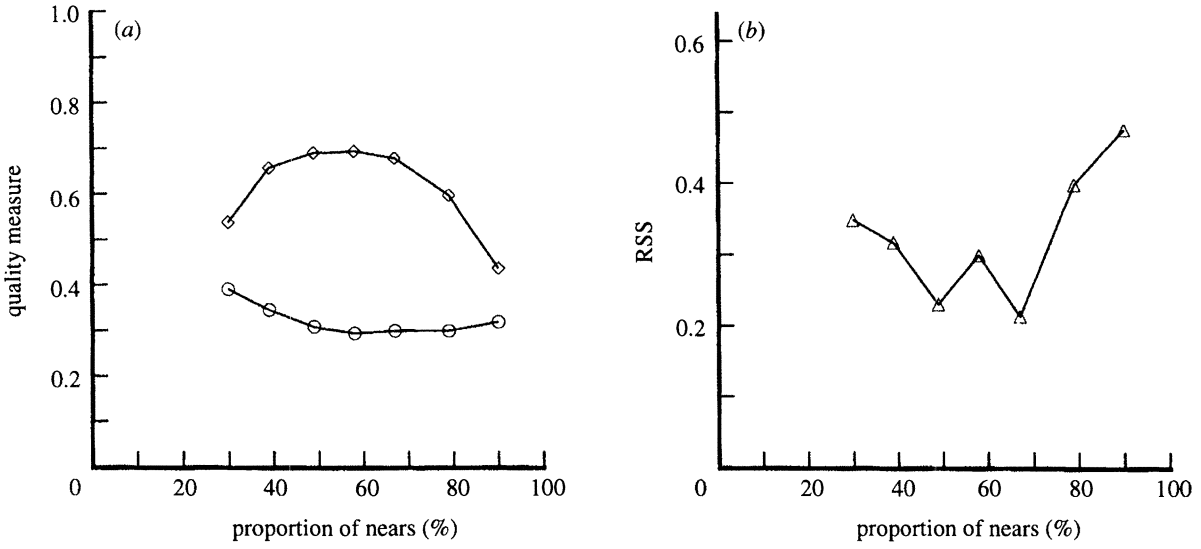

Figure 3. Graphs showing various measures of the quality of the derived city distributions as a function of P_N. (*a*) RSQ (diamonds) and SSTRESS (circles). (*b*) RSS.

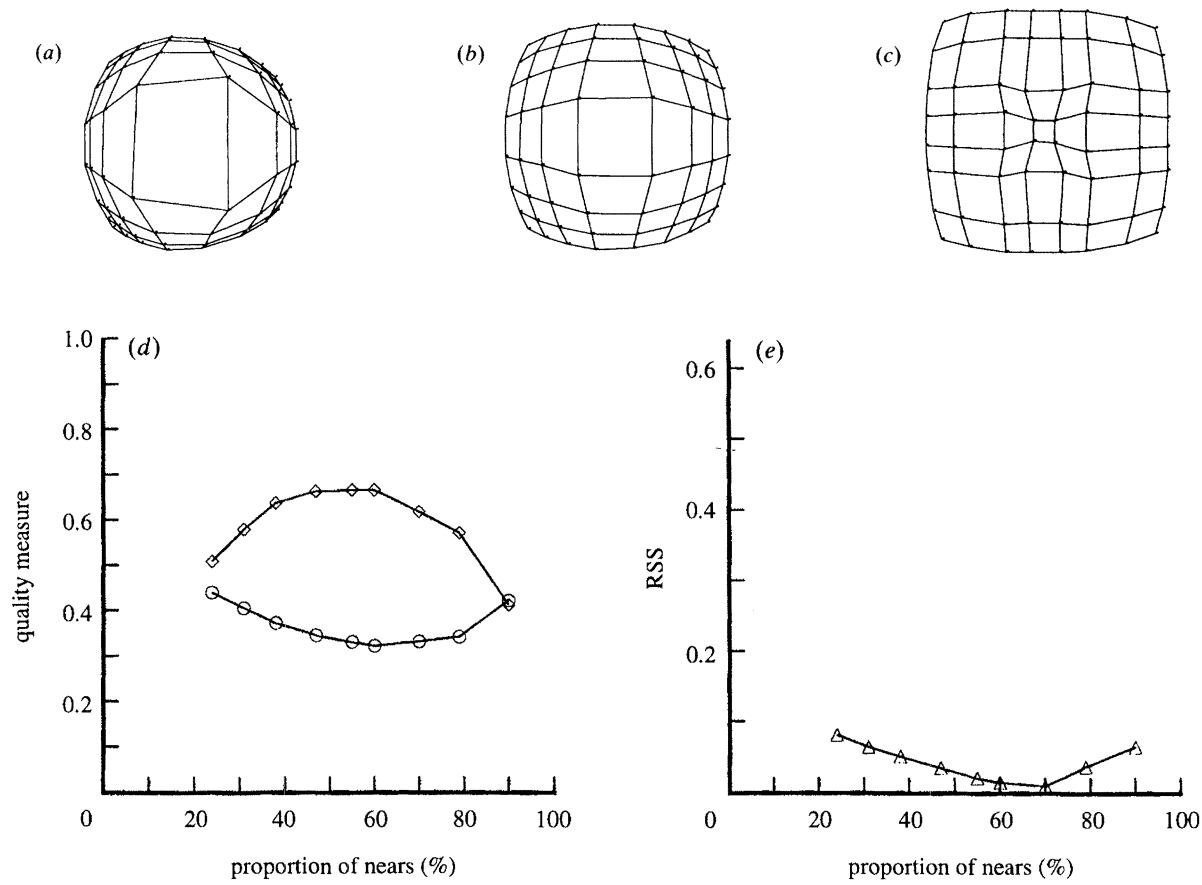

Figure 4 (*a–c*). Configurations computed from the grid distribution by NMDS. P_N = 25 %, 50 % and 75 % respectively. Points that are adjacent to each other in the grid underlying the dissimilarity matrix are connected by lines. (*d, e*). Graphs showing various measures of the quality as a function of P_N of the configurations computed for the grid example. (*d*) RSQ (diamonds) and SSTRESS (circles). (*e*) RSS.

Three different numbers of points were investigated: $N = 15$, $N = 30$, and $N = 45$, with 20 different configurations for each value of N. Binary matrices were generated as described in the previous section, P_N ranging from 0.2 to 0.9 in increments of 0.1.

Figure 5 shows that the SSTRESS, RSQ and RSS curves show the same variation with P_N as found in the cities and grid examples. RSS decreases with N, as expected from metric determinacy arguments. A different trend is indicated by the apparent fit measures: e.g. RSQ is consistently higher (i.e. better) for $N = 15$ than for $N = 30$. Thus increasing N aids accurate recovery but this is not reflected in the apparent fit measures, which usually are all that is available. This unwelcome effect is not specific to binary data: Kruskal & Wish (1978) noted that STRESS values tend to grow with increasing

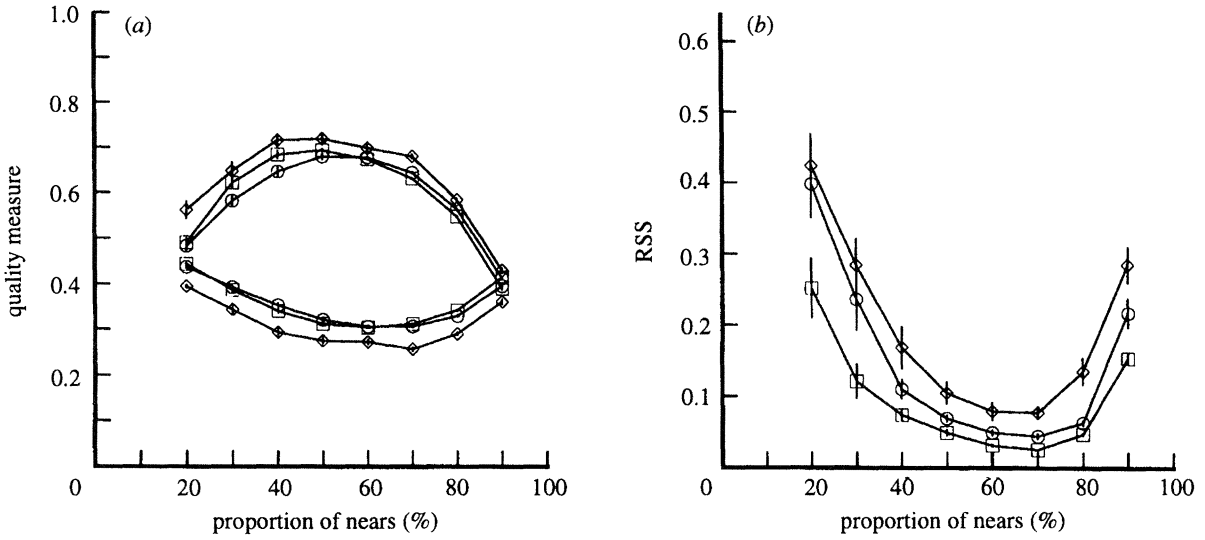

Figure 5. Performance measures for Monte Carlo runs as a function of P_N, for different numbers of configuration points N (diamonds, $N = 15$; circles, $N = 30$; squares, $N = 45$). Vertical bars are standard errors over 20 runs. (*a*) Mean RSQ (upper curves) and SSTRESS (lower curves). (*b*) Mean RSS.

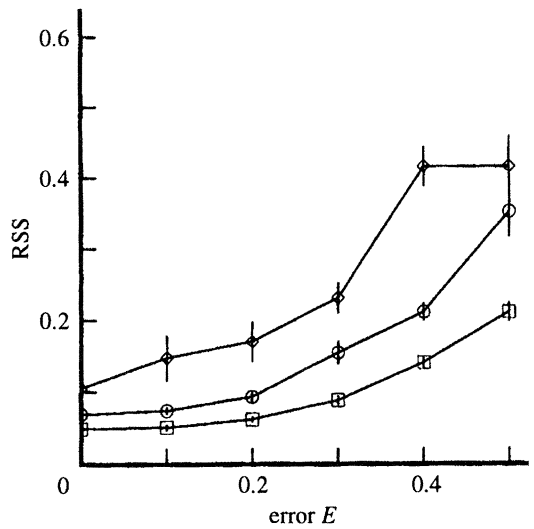

Figure 6. Values of mean RSS for Monte Carlo runs as error is added into the process of generating the dissimilarities, for $P_N = 50\,\%$ (diamonds, $N = 15$; circles, $N = 30$; squares, $N = 45$). Vertical bars are standard errors over 20 runs.

N for small values of N, but become stable for larger values of N.

Real dissimilarity measurements are almost always prone to error, and we wished to test the influence of this in the binary case. Error was introduced into the full 'distances' d_{ij} by the following standard formula (Young 1970):

$$d_{ij} = \left[\sum_{k=1}^{2} (x_{ik} + e_{ijk} - x_{jk} - e_{jik})^2 \right]^{\frac{1}{2}},$$

where x_{ik} is the kth coordinate of point i and e_{ijk} is an independent random variable drawn from a gaussian distribution. The standard deviation σ_E of the gaussian is defined in terms of an error level E through $\sigma_E = E\sigma_c$, where σ_c is the standard deviation of the coordinates in the particular configuration. Results for $P_N = 50\,\%$ are shown in figure 6. Performance decreases as the added error increases, with greater rapidity for small N than large N. For further discussion see Simmen (1994).

(*d*) *Annular structures*

Here we show that when ALSCAL is used to scale binary data using the tied approach, in two situations the derived configuration has an *annular* structure. These situations are when (i) the data has low P_N or (ii) the derived configurations have low fit.

Figures 2 and 4 show that scaling the city and grid distributions gives a marked annularity to the derived configuration for $P_N \leqslant 50\,\%$, with the strongest effect at low P_N. We explored this effect systematically by examining the configurations produced in the Monte Carlo experiments. The dissimilarity matrices used in those experiments were derived from disc configurations. Each plot of figure 7 is the superposition of the solutions from the 20 different runs for $N = 30$ points, at given values of P_N and added error, E. For no added error, shown in the top row, there is again a strong tendency towards annular solutions for low P_N. The plot in which the points are spread most uniformly over the disc is that for $P_N = 70\,\%$; this value of P_N also gives the best true fit (mean RSS = 0.044; see figure 5). Equivalent superposition plots for $N = 15$ and $N = 45$ (not presented here) show similar trends.

The second bias towards annularity can arise in derived configurations with poor fit. The middle and bottom rows of figure 7 illustrate this, showing changes in the overall shape of the solutions as the dissimilarity matrix is corrupted by error, leading to lower fit. Quantitative measures of performance for the three error levels are shown in figure 8.

If ALSCAL has a tendency to produce annular configurations for low P_N or in cases of low fit, then in such cases reconstructions of configurations that have *true* annular structure should be better than reconstructions of disc configurations. To test this, the Monte Carlo study described in §3*c*(ii) was repeated,

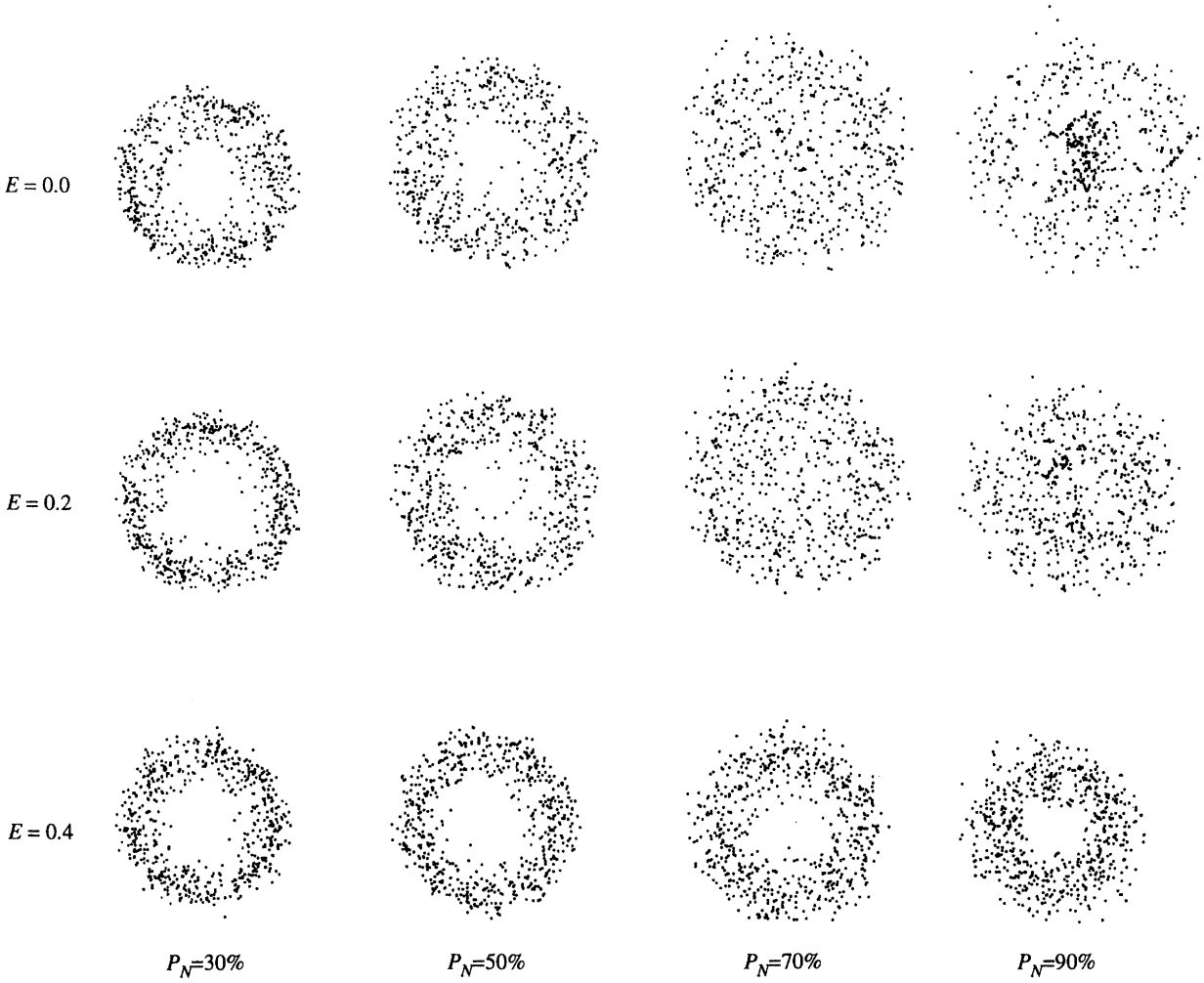

Figure 7. The solutions obtained by scaling binary dissimilarity matrices derived from configurations of 30 points taken randomly from a disc. Each plot shows the superposition of 20 solutions, drawn about their common centroid. Results are shown for different values of added error, E, and the proportion of 'nears', P_N. Note the similarity of the trend shown in the top row to that seen in figure 4 (*a–c*).

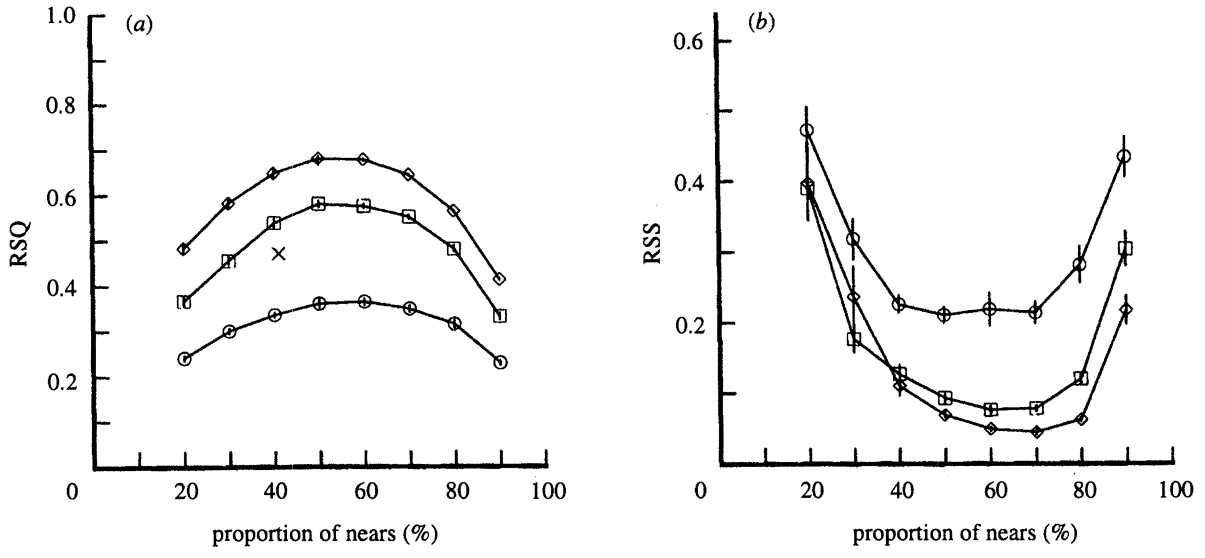

Figure 8. Performance for disc distributions with $N = 30$, for various levels of additional error (diamonds, $E = 0.0$; squares, $E = 0.2$; circles, $E = 0.4$). Each point is the average computed from 20 runs, vertical bars indicate standard errors. ' × ' marks the data point 'vs', explained in §4 in the text. (*a*) RSQ. (*b*) RSS.

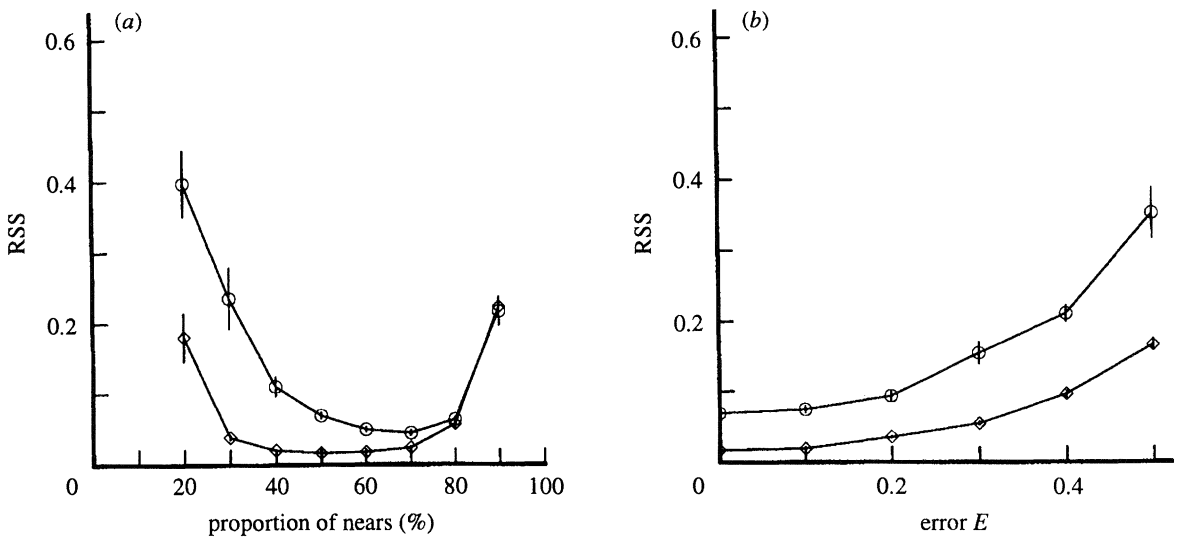

Figure 9. True fit (RSS) for the reconstructions of annular distributions (diamonds) compared to the fit for uniform distributions (circles). $N = 30$. Each point is the average computed from 20 runs, vertical bars indicate standard errors. (*a*) Variation with P_N, no added error. (*b*) Variation with error level for $P_N = 50\%$.

but with points drawn randomly from the annulus with internal radius 0.7 and external radius 1.0. As shown in figure 9*a*, for low P_N reconstructions of annular configurations are substantially more accurate than those of disc configurations, the effect diminishing as P_N increases. Figure 9*b* shows that as more error is added into the dissimilarity matrices to lower the fit, the difference in RSS between reconstructed annular and disc configurations increases.

(i) *Why do annular structures occur?*

The bias towards the production of annular structures for low P_N data is due to the constraints imposed in the tied approach. Suppose that data from a square grid of points is scaled and that the near/far distance threshold is chosen such that for a point close to the centre of the grid only its four neighbours are regarded as 'near'. Consider the value of SSTRESS calculated for a perfect reconstruction of the grid. Although the constraint that such a point be close to and equidistant from its 'near' points is satisfied, giving no contribution to SSTRESS, there is a broad distribution in the distances from it to the large number of 'far' points, giving a large contribution to SSTRESS. If now all points are pushed out towards the edge (as in figure 4*a*), there will be a small contribution to SSTRESS from the population of 'near' points. There will be a much larger reduction in SSTRESS from the population of 'far' points, since there is now a smaller spread in their distances from the reference point.†

The bias towards annularity produced by low fit was first conjectured for SSTRESS by de Leeuw & Bettonvil (1986). From a mathematical analysis they concluded that 'especially in the case of poor fit, multidimensional scaling solutions based on SSTRESS may be biased towards distributing clusters of points regularly over the surface of a sphere.' In further support of this, we note that extremely poor fit solutions can be generated by scaling random data with no underlying structure (equivalent to $E \rightarrow \infty$ in our protocol) and that these are strongly annular in form, irrespective of whether the dissimilarities are binary or not (our unpublished observations).

(ii) *Seriation*

The generation of curved structures by NMDS (and other methods) has been noted in the context of the problem of seriation (Kendall 1971*b*), which is concerned with the extraction of an underlying one-dimensional order. In archaeology, for example, it is often desired to find the temporal ordering of a number of grave sites, given data on the incidence of a number of distinct artefacts at each site and the assumption that each type of artefact was produced in just one period of history. The dissimilarity between any two graves is calculated as a function of the overlap between the collections of the artefacts they contain. NMDS is then applied using the tied approach, standardly in two dimensions. Given suitable data, a roughly one-dimensional configuration can be obtained, yielding the appropriate serial order. Rather than the entities lying on a straight line in the two-dimensional NMDS plot, they are usually bent into a shape conventionally termed a horseshoe (see, for example, Kendall 1971*b*, 1975; Shepard 1974; Gower 1987). For the purposes of seriation this effect does not hinder interpretation, provided the tips of the horseshoe are readily identifiable.

As has been discussed by Kendall (1971*b*, 1975), among others, horseshoes can arise from the constraints imposed in the tied approach. Consider the case where the dissimilarity between any two entities is low for entities nearby in the sequence, increases with increasing separation and then reaches a maximum.

† For zero error and large P_N, figure 7 also shows that points cluster in the centre of the configurations. There are many 'near' distances and few 'fars' and so to minimize the SSTRESS the distribution of distances associated with 'nears' must be as narrow as possible, the distribution of 'fars' being of less importance. Note that this trend is abolished for low fit.

If in the derived configuration the entities are arranged in the correct order and in a straight line, there will be a large spread in the distances associated with the maximal dissimilarities and hence a large contribution to SSTRESS; it is more favourable to place the entities around a horseshoe. This effect is most marked when many dissimilarities have maximum value, which for the binary case corresponds to low P_N.

The shape of the horseshoe will depend on the objective function. STRESS is a quadratic function of the discrepancies between distances and disparities, and thus its value is influenced more by large distances than by small ones. This effect is even stronger for SSTRESS, which contains quartic terms. Thus ALSCAL (which minimizes SSTRESS) would be expected to have a strong tendency to produce horseshoes. Other non-NMDS techniques for seriation also tend to produce curved structures. In ecology, for example, controversy exists about the use of detrended correspondence analysis (Hill & Gauch 1980) to remove the horseshoe effect, with researchers disputing whether or not the effect reflects genuine structure: see for example Wartenberg *et al.* (1987) and the subsequent correspondence.

4. RE-EXAMINING THE VISUAL SYSTEM CONFIGURATION

In §2 we derived an NMDS configuration for 30 areas of the primate cortical visual system (figure 1). Using our Monte Carlo results for configurations of 30 points, we now assess whether this configuration, which we henceforth refer to as vs, is likely to be biased towards annularity by either low P_N or low fit. The P_N value for our similarity matrix derived from the visual system data is 41%. From figure 7, annularity is apparent in reconstructions of disc configurations for both $P_N = 30\%$ and $P_N = 50\%$, even for zero added error. Regarding the fit, vs has an RSQ value of 0.47, as compared to the RSQ ceiling value (constraining P_N to equal 41%) of 0.73 (effectively the 'perfect' fit value using the tied approach). Scrutiny of figure 8*a* shows that this low fit value is comparable with that obtained by scaling data from disc distributions with a noise level $E \approx 0.3$, for a similar value of P_N. This equivalence holds not just for the RSQ values but also for these normalized by the RSQ ceiling values: the mean ceilings (fixing P_N to 40%) for the Monte Carlo results for $N = 30$ and $P_N = 40\%$ are 0.74 ± 0.03 and 0.72 ± 0.02 for the $E = 0.2$, 0.4 cases respectively. Figure 7 indicates that the reconstructions in the $E \approx 0.3$ case would be markedly annular.

We have conducted additional Monte Carlo tests that confirm that these findings still hold in the case of ternary level similarity data, as was used by Young (1992). For example, the experiments of §3*c* (ii) on scaling data derived from disc distributions were repeated using ternary level data. The proportions of 'high', 'medium', and 'low' similarity elements were chosen to match those in the ternary similarity matrix derived from the visual system FVE matrix according to Young's assignment rules (see section §2). Figure 10*a* shows that the reconstructions again have an annular form, whereas figure 10*b* shows that the reconstructions from $E = 0.3$ data have RSQ values similar to that obtained scaling the ternary similarity matrix for the visual system.

On the basis of these comparisons, we suggest that the NMDS configuration derived from the visual system

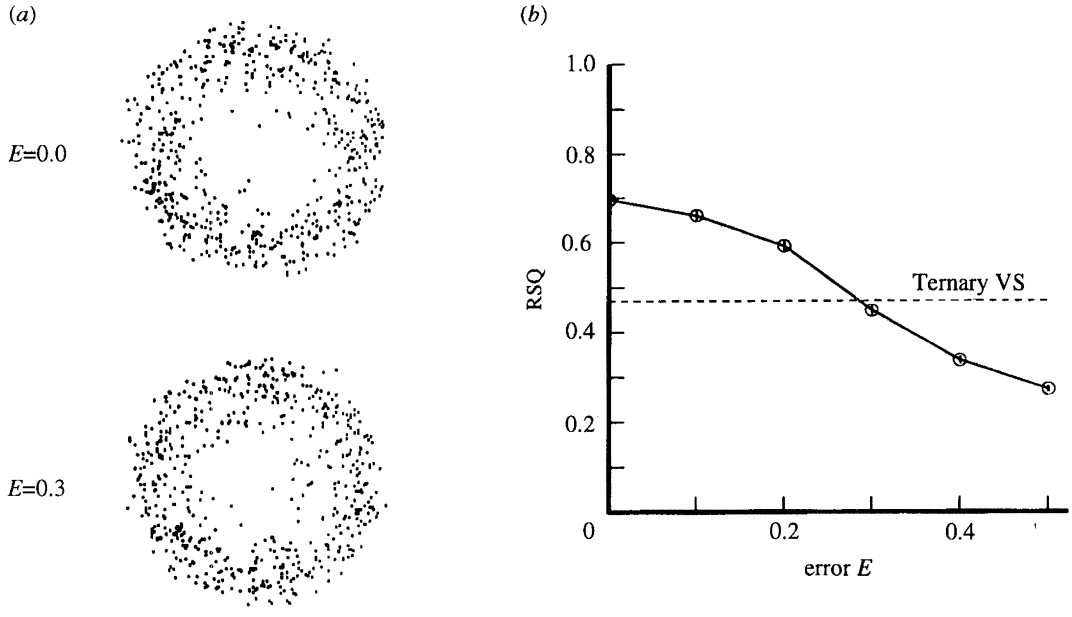

Figure 10. Results from scaling ternary similarity data derived from the same configurations of 30 points taken randomly from a disc that were used in §3*c* (ii), using a simple generalization of the procedure used to produce binary data. The numbers of similarity values assigned values 0, 1, and 2 were fixed to be 255, 57, and 123, these being the equivalent numbers in the ternary similarity matrix derived from the visual system FVE matrix. (*a*) Superposition plots of the 20 solutions, drawn about their common centroid, for two levels of added error in the similarity data. (*b*) The mean RSQ value as a function of the level of added error. Vertical bars are standard errors computed over 20 runs. The dashed line indicates the value of RSQ (0.47) obtained scaling the ternary similarity matrix derived from the FVE matrix.

(*a*)

1												
2	1											
3	1	1										
4	0	1	1									
5	0	0	1	1								
6	0	0	0	1	1							
7	0	0	0	0	1	1						
8	0	0	0	0	0	1	1					
9	0	0	0	0	0	0	1	1				
10	0	0	0	0	0	0	0	1	1			
11	0	0	0	0	0	0	0	0	1	1		
12	0	0	0	0	0	0	0	0	0	1	1	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**

(*b*)

1												
2	1											
3	1	1										
4	0	1	1									
5	0	0	1	1								
6	0	0	0	1	1							
7	0	0	0	0	1	1						
8	0	0	0	0	0	1	1					
9	0	0	0	0	0	0	1	1				
10	0	0	0	0	0	0	0	1	1			
11	1	0	0	0	0	0	0	0	1	1		
12	1	1	0	0	0	0	0	0	0	1	1	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**

Figure 11. Example binary similarity matrices displaying serial order for 12 entities. (*a*) Serial order with distinct endpoints. (*b*) Cyclical serial order. These two forms are closely related to what are termed simplex and circumplex structures respectively in the psychological literature (Guttman 1954; Shepard 1974). Diagonal elements (which by convention have value 1) are not shown.

data is corrupted by artefact. NMDS has been applied to connectivity datasets from other regions of the brain, and most of the configurations published also have low RSQ and a strongly annular or horseshoe shape (Young 1992, 1993; Scannell & Young 1993). In the absence of independent corroboration it is most reasonable to assume that these configurations also have an artefactual component.

(*a*) *An independent analysis based on seriation*

We have used a graph-theoretic method derived from the seriation literature to attempt a corroboration for the visual system data. We discuss how far it is possible to order the set of entities so that each entity has a similarity of 1 (or dissimilarity of 0) to the entities closest to it in the sequence and a similarity of 0 (dissimilarity of 1) to all other entities. We distinguish orderings in which the sequence has two distinct endpoints from those where the sequence is cyclical. Figure 11 shows similarity matrices corresponding to these cases, the order of rows (or columns) specifying the ordering of entities. In the case of cyclical order, in addition to the diagonal band of 1s there is a cluster at the off-diagonal corners. Assessing whether the entities can be arranged in a serial order is equivalent to permuting the rows and columns of the symmetrical similarity matrix to see if it can be cast into one of these characteristic forms.

A measure of the degree to which these ideal matrix forms can be approached is provided if each entity is regarded as a point in N-dimensional space, with coordinates given by the elements in the corresponding row of the similarity matrix and distances given by the Hamming measure. Expressed in these terms, the best permutation of rows and columns for the cyclical case is that which specifies the shortest closed circuit around the N points, also known as the solution to the Travelling Salesman Problem (Lawler *et al.* 1985).† It is straightforward to show that the circuit length for perfect ordering is $2N$; lengths greater than this indicate imperfect ordering. Analogous considerations hold for non-cyclical sequences. If the annular form of the visual system NMDS solution VC strongly reflects an ordering of the visual areas into two streams sharing a common origin and destination, i.e. a cyclical ordering, then the shortest circuit length for the similarity matrix used to produce vs should be near 60. Using a modified 3-opt algorithm (Lawler *et al.* 1985) to find short circuits, the best circuit length we obtained (over 30000 runs) was 204.

Interpreting quantitative measures of imperfect serial orderings is known to be difficult (Hubert 1974). We therefore compared the visual system result with those from Monte Carlo studies using geometrical configurations. Table 1 shows the minimal circuit lengths obtained for binary similarity matrices derived in §3 from the disc and annulus configurations. Using similarity data derived by thresholding exact distances between points distributed randomly in an annulus produces minimum circuit lengths close to the theoretical minimum. As noise is introduced into the similarity data so the ordering becomes less perfect and the circuit lengths grow. Circuit lengths for data derived from configurations with points lying randomly in a disc serve as a control, representing the results expected when there is no underlying serial order to the entities. The minimal circuit lengths in this case are always higher than in the annular case (though as the noise in the similarity data increases, the lengths show signs of converging). The disc data in table 1 confirms that the increasing annularity of NMDS solutions as the apparent fit level decreases (illustrated by figure 7) does not reflect genuine ordering.

Comparing the visual system result of 204 with the table 1 data, we note that 204 is far in excess of 107.2, the mean minimal circuit length for data-sets derived with no added noise from disc-like configurations (i.e. data-sets possessing no obvious serial order). Furthermore, the visual system result is consistent with those from similarity matrices generated from disc-like configurations with a considerable degree of noise, specifically, those with $E = 0.3$. This is a striking result, since 0.3 is also the level of noise required in the Monte Carlo

† Wilkinson (1971) and Hubert (1974) have discussed how the related task of seriating entities given an entity-by-attribute matrix is also equivalent to a shortest circuit problem.

Table 1. *Minimal circuit length* L_{min} *as a function of the level of error* (E) *incorporated in similarity matrices derived from genuine disc and annular configurations with* $N = 30$ *and* $P_N = 40\%$

(Each average was computed over 20 instances, each defined on a similarity matrix from the studies of §§3*c*(ii) and 3*d*. For each instance, 500 runs (with different initial circuits) were performed with a variant of the 3-opt algorithm, and the minimum circuit length recorded. The average length of a random circuit on a random binary similarity matrix is $2P_N(100-P_N)N^2 \times 10^{-4}$: for $N = 30$ and $P_N = 40\%$ this equals 432.)

	disc	annulus
E	mean ± s.d. of L_{min}	mean ± s.d. of L_{min}
0.0	107.2 ± 11.4	61.2 ± 1.5
0.1	125.9 ± 13.0	79.2 ± 4.3
0.2	164.5 ± 13.0	118.9 ± 9.7
0.3	213.5 ± 14.3	167.6 ± 13.6
0.4	242.0 ± 10.2	220.2 ± 14.1
0.5	265.0 ± 11.6	256.6 ± 12.8

runs to give an RSQ value matching that for vs. Thus seriation and NMDS, though independent methods, give consistent results.

5. DISCUSSION

The areas found in the two arms of the annulus in vs are roughly those found in the temporal and parietal lobes, in agreement with the two streams hypothesis. However, the tendency towards annularity in the NMDS approach suggests that the existence of other types of structure in the data would be obscured by this analysis. Despite the appearance of vs, on the evidence of the P_N and RSQ values one interpretation could be that the visual areas have an underlying abstract connectivity structure in which they are positioned throughout a circular region, reflecting little organization into streams. NMDS will find annular structure whether it exists in the data or not. Various authors (e.g. Martin 1988; Merigan & Maunsell 1993) have argued on biological grounds that two streams is a rough approximation to the truth, but that this view obscures the existence of other types of structure.

(*a*) *Counterarguments to the claim of artefactual structure*

We now address the various counterarguments given by Young *et al.* (1994) to the case presented in this paper and in (Simmen *et al.* 1994). Firstly, they state that their results in press show that good (low RSS) solutions can be obtained from dissimilarities 'at the same level of measurement as connectivity data' derived from geometric configurations. Secondly, they state that artefactual structure only arises for RSQ values significantly lower than that obtained for the visual system data. The results presented in the current paper contradict both these assertions (see for instance figures 7, 8 and 10). Thirdly, using our seriation-based method, they computed the circuit length for their similarity matrix, and emphasized that the length of their circuit lies well outside the distribution for random circuits. We do not dispute this; however, a more relevant point is that circuits shorter than those derived from the visual system data can be produced from data with no underlying serial order (see table 1). Fourthly, they derived NMDS configurations for the visual system data in up to six dimensions, and claim that these still possess a near-planar circular form (though they do not present quantitative results). To investigate this, we submitted our binary similarity matrix for the visual system data to ALSCAL in three dimensions, and then calculated the eigenvalues for the three principal components of the resulting configuration. These measure the variance in three directions, with the third component giving the variance in the 'flattest' direction. If the configuration were a near-planar circular ring, one would expect the first two components to be much larger than the third component. We obtained values for the three components of 1.36, 0.93 and 0.81, indicating little flattening of the configuration into a plane. The equivalent values using Young's tenary matrix are 1.37, 0.94 and 0.79.

(*b*) *Alternative NMDS approaches*

We have identified various problems in interpreting NMDS representations of dissimilarity matrices possessing few distinct levels. We now discuss whether these problems can be lessened by adopting either of two standard variations in the way NMDS is applied. These are (i) using the untied rather than tied approach, and (ii) using a 'higher-order' dissimilarity matrix.

(i) *The untied approach*

Takane *et al.* (1977) argued that the untied approach is appropriate if the data are drawn from a continuous distribution and ties are due to rounding errors introduced by the measurement process, and that the tied approach is appropriate if the data are drawn from an inherently discrete distribution. Connections between brain areas will be of different strengths. However, the data in the FVE matrix is specified only in terms of the presence or absence of connections, reflecting the difficulty of ascribing more quantitative values from the currently available biological data. If one labels pathways as having one of several strengths, then this implies a discretization of a presumably continuous underlying distribution. The arguments of Takane *et al.* (1977) indicate that the untied approach should be used in this case. However, in the binary case, one could regard presence/absence as being an inherently discrete distribution, in which case the tied approach is more appropriate. Thus the tied/untied issue is much clearer for the case of many levels than the binary case.

A practical problem of using the untied approach with binary data is that the apparent fit measures always indicate near-perfect reconstruction, irrespective of the true fit (our unpublished observations). This

is simply because many fewer constraints are required to be satisfied than in the tied case. For further discussion see Coxon (1982, p. 53).

For comparison purposes we derived the untied configuration for our binary similarity matrix for the visual system data. The RSQ was 0.98, and areas were spread more uniformly throughout the region than in figure 1. However, given the small number of constraints, it is possible to move areas by quite large amounts while barely changing the apparent fit. Thus though adopting the untied approach ameliorates the annular bias, it also introduces problems of its own.

(ii) *Higher order dissimilarity measures*

MDS works on an $N \times N$ matrix of dissimilarities between entities. Sometimes the raw data is an $N \times M$ entity-by-attribute matrix, each row giving measurements of M attributes for a single entity. To apply MDS to such data a dissimilarity matrix must be generated by calculating dissimilarities between the attribute vectors (Sneath & Sokal 1973; Gower 1971*a*; Krzanowski 1988). It is always possible to regard a dissimilarity matrix as a particular type of entity-by-attribute matrix in which the entries are themselves dissimilarities, and then derive a higher-order dissimilarity matrix by the same methods (Kruskal & Wish 1978; Rosenberg & Jones 1972). As the new matrix generally has more distinct data values than the original one, this approach might appear useful for overcoming the problem of ties in poorly differentiated data-sets. However, Monte Carlo studies show that NMDS solutions obtained using higher-order dissimilarities either do not give improved recovery of structure (Drasgow & Jones 1979), or – for binary data – do so only for a particular row-comparison measure for data with low P_N (Simmen 1994), a regime in which the absolute levels of recovery are in any case rather poor. However, there is a rather different type of derivation rule relevant only to binary data.

This rule is easily understood when the raw dissimilarities convey neighbourhood information. One of the first examples was abuttal data for 88 of the departments of France (Kendall 1971*a*). Each department abuts at most eight others, so P_N is somewhat removed than 0.1. The results we presented above suggest that reconstruction using the abuttal matrix directly would be extremely poor. However, Kendall derived a new dissimilarity matrix by setting the dissimilarity to be the smallest number of departments stepped through to join each pair. There are at least a dozen levels in the new matrix, and the map of France obtained by scaling it is very good (Kendall 1971*a*). This stepping algorithm can be applied to any binary dissimilarity matrix. However, it will work well only in certain circumstances.

A map of France drawn on an elastic sheet can be arbitrarily deformed without changing the abuttal data. So why is Kendall's map realistic? The departments used are roughly the same size, and are laid out roughly isotropically (the small Parisian departments were excluded). In this case, the stepping dissimilarity approximates the Euclidean distance: discretized into several levels. In contrast, consider the mainland states of the USA. Good reconstruction can be obtained for the western states. However, as the eastern states are added, which are much less regular in shape and size, reconstruction deteriorates markedly (our unpublished observations).

The degree to which the entities are the same 'size' and uniformly distributed will determine the quality of the reconstruction using the stepping rule matrix. Similar arguments have been made by McGinley (1977) and Sibson *et al.* (1981). The stepping rule matrix for our similarity matrix derived from the visual system data contains only three levels, and thus offers little improvement over the original binary matrix.

(*c*) *Higher levels of detail*

The data on which this analysis is based so far ignores the fact that many of the areas in the visual system are subdivided into regions or laminae with different functional properties. In V1, for example, the processing of colour and motion appear to be segregated (Zeki & Shipp 1988). It may be more appropriate to treat these different sub-regions as distinct entities for the purposes of an NMDS analysis. This type of analysis also ignores information available about the hierarchical relation between areas. There is a very stereotyped pattern of laminar origin and termination for connections between areas, depending on whether the areas are at the same level in the processing hierarchy, or whether one is above or below the other (Felleman & Van Essen 1991). Such information can be utilized to produce wiring diagrams incorporating hierarchial structure based on alternative principles to those of NMDS (Felleman & Van Essen 1991, figure 4).

6. CONCLUSIONS

We have: (i) investigated the usefulness of NMDS for reconstructing known configurations from binary data; (ii) shown that two separate biases can cause artefactual, annular structure in the derived configurations; and (iii) used these findings to address the reliability of applications of NMDS to connectivity data in the brain.

1. Reconstruction of known configurations from binary dissimilarity data can vary, the quality improving as the number of points (and thus the number of constraints) increases. Another factor affecting quality is P_N, the proportion of nears in the dissimilarity matrix. The best reconstruction is obtained for P_N in the range 50–60%; for $P_N \geqslant 80\%$ or $P_N \leqslant 20\%$, reconstruction is very poor. A more important factor affecting reconstruction quality is the nature of the true configuration. Reconstruction of configurations of points arranged in an annulus, for example, is much better than reconstructions of points distributed uniformly within a circle.

2. Artefactually annular NMDS configurations can arise for low P_N and for low fit. Our results indicate how measures of apparent fit might be used to predict the circumstances under which the bias due to low fit

is likely to play a role. These results also explain why reconstruction is better when the true configuration is annular. More generally, the results of any application of NMDS to dissimilarity data containing a low number of levels should be treated with caution.

3. The application of NMDS to connectivity data for the visual system results in a configuration which is strongly annular. It is highly likely that a component of the annularity is due to artefact. This considerably weakens the support that this configuration can provide for the two streams view of visual processing. Similar conclusions apply to the other recent applications of NMDS to connectivity data (Young 1993; Scannell & Young 1993). Annular NMDS configurations may be consistent with many interpretations of the underlying structure of the data.

We are very grateful to Bruce Graham, David Price and Nick Chater, and the four anonymous referees, for their helpful comments. Financial support is acknowledged to the Joint Councils Initiative for Cognitive Science/HCI (G.J.G.), to the MRC Research Centre for Brain and Behaviour at Oxford (M.W.S.) and to the MRC under Programme Grant no. 9119632 (D.J.W.).

REFERENCES

Baizer, J.S., Ungerleider, L.G. & Desimone, R. 1991 Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J. Neurosci.* **11**, 168–190.

Bennett, J.F. & Hays, W.L. 1960 Multidimensional unfolding: determining the dimensionality of ranked preference data. *Psychometrika* **25**, 27–43.

Coxon, A.P.M. (ed.) 1982 *The user's guide to multidimensional scaling*. London: Heinemann Educational Books.

de Leeuw, J. & Bettonvil, B. 1986 An upper bound for sstress. *Psychometrika* **51**, 149–153.

Desimone, R. & Ungerleider, L.G. 1989 Neural mechanisms of visual processing in monkeys. In *Handbook of neuropsychology*, vol. 2 (ed. F. Boller & J. Grafman), pp. 267–299. Amsterdam: Elsevier Science.

DeYoe, E.A. & Van Essen, D.C. 1988 Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* **11**, 219–226.

Digby, P.G.N. & Kempton, R.A. 1987 *Multivariate analysis of ecological communities*. London: Chapman and Hall.

Drasgow, F. & Jones, L.E. 1979 Multidimensional scaling of derived dissimilarities. *Multivar. Behav. Res.* **14**, 227–244.

Durbin, R. & Mitchison, G. 1990 A dimension reduction framework for understanding cortical maps. *Nature, Lond.* **343**, 644–647.

Durbin, R. & Willshaw, D.J. 1987 An analogue approach to the travelling salesman problem using an elastic net method. *Nature, Lond.* **326**, 689–691.

Felleman, D.J. & Van Essen, D.C. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1**, 1–47.

Goodale, M.A. & Milner, D.A. 1992 Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25.

Gower, J.C. 1966 Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Gower, J.C. 1971*a* A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857–874.

Gower, J.C. 1971*b* Statistical methods of comparing different multivariate analyses of the same data. In *Mathematics in the archaeological and historical sciences* (ed. F. R. Hodson, D. G. Kendall & P. Tăutu), pp. 138–149. Edinburgh University Press.

Gower, J.C. 1982 Discussion following Ramsey (1982).

Gower, J.C. 1987 Introduction to ordination techniques. In *Developments in numerical ecology* (ed. P. Legendre & L. Legendre), pp. 3–64. Berlin: Springer Verlag.

Greenacre, M.J. & Underhill, L.G. 1982 Scaling a data matrix in a low-dimensional euclidean space. In *Topics in applied multivariate analysis* (ed. D. M. Hawkins). Cambridge University Press.

Guttman, L. 1954 A new approach to factor analysis: the radex. In *Mathematical thinking in the social sciences* (ed. P. F. Lazarsfeld), pp. 258–348. Glencoe, Illinois: Free Press.

Hasselmo, M.E., Rolls, E.T., & Baylis, G.C. 1989 The role of expression and identity in the face-selective responses of neurons in the temporal visual-cortex of the monkey. *Behav. Brain Res.* **32**, 203–218.

Hess, S.T., Blake, J.D. & Blake, R.D. 1994 Wide variations in neighbour-dependent substitution rates. *J. molec. Biol.* **236**, 1022–1033.

Higgins, D.G. 1992 Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets. *Comput. Applic. Biosci.* **8**, 15–22.

Hill, M.O. & Gauch, H.G. Jr 1980 Detrended correspondence analysis: an improved ordination technique. *Vegetatio* **42**, 47–58.

Hubert, L. 1974 Some applications of graph theory and related non-metric techniques to problems of approximate seriation: the case of symmetric proximity measures. *Br. J. math. statist. Psychol.* **27**, 133–153.

Jensen, R.J. & Barbour, C.D. 1981 A phylogenetic reconstruction of the Mexican cyprinid fish genus *Algansea*. *Syst. Zool.* **30**, 41–57.

Kendall, D.G. 1971*a* Construction of maps from "odd bits of information". *Nature, Lond.* **231**, 158–159.

Kendall, D.G. 1971*b* Seriation from abundance matrices. In *Mathematics in the archaeological and historical sciences* (ed. F. R. Hodson, D. G. Kendall & P. Tăutu), pp. 215–252. Edinburgh University Press.

Kendall, D.G. 1975 The recovery of structure from fragmentary information. *Phil. Trans. R. Soc. Lond.* A **279**, 547–582.

Klahr, D.A. 1969 Monte Carlo investigation of the statistical significance of Kruskal's scaling procedure. *Psychometrika* **34**, 319–330.

Kohonen, T. 1982 Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69.

Kruskal, J.B. 1964*a* Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27.

Kruskal, J.B. 1964*b* Non-metric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129.

Kruskal, J.B. & Wish, M. 1978 *Multidimensional scaling*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07–011. Beverly Hills, California: Sage Publications.

Krzanowski, W.J. 1988 *Principles of multivariate analysis: a user's perspective*. Oxford Statistical Science Series. Oxford University Press.

Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G. & Shmoys, D.B. 1985 *The traveling salesman problem*. Chichester: Wiley.

Legendre, P. & Legendre, L. 1987 *Developments in numerical ecology*. Berlin: Springer Verlag.

Livingstone, M. & Hubel, D. 1988 Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, Wash.* **240**, 740–749.

Lowe, D. 1993 Novel topographic nonlinear feature extraction using radial basis functions for concentration coding in the artificial nose. *Proceedings of the Third International Conference on Artificial Neural Networks*, pp. 95–99. London: IEE.

Martin, K.A.C. 1988 From enzymes to visual perception: a bridge too far?. *Trends Neurosci.* **11**, 380–387.

Maunsell, J.H.R. & Newsome, W.T. 1987 Visual processing in monkey extrastriate cortex. *A. Rev. Neurosci.* **10**, 363–401.

McGinley, W.G. 1977 Some optimisation problems in data analysis. Ph.D. thesis, University of Cambridge.

Merigan, W.H. & Maunsell, J.H.R. 1993 How parallel are the primate visual pathways?. *A. Rev. Neurosci.* **16**, 369–402.

Mishkin, M., Ungerleider, L.G., & Macko, K.A. 1983 Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417.

NAG 1991 *Fortran Library, Mark* 15. Oxford: Numerical Algorithms Group.

O'Mara, S.M., Scannell, J.W., Burns, G., & Young, M.P. 1994 A topological analysis of the connectivity of the hippocampal formation. *Brain Research Association Meeting* 1994 *Abstracts, vol.* 11, *p.* 65.

Ramsay, J.O. 1982 Some statistical approaches to multidimensional scaling data (with discussion). *Jl R. statist. Soc.* A **145**, 285–312.

Rohlf, F.J. 1970 Adaptive hierarchical clustering schemes. *Syst. Zool.* **19**, 58–82.

Rosenberg, S. & Jones, R.A. 1972 A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *J. Personality Social Psychol.* **22**, 372–386.

Scannell, J.W. & Young, M.P. 1993 The connectional organization of neural systems in the cat cerebral cortex. *Curr. Biol.* **3**, 191–200.

Schönemann, P.H. & Carroll, R.M. 1970 Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* **35**, 245–255.

Shepard, R.N. 1962*a* The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140.

Shepard, R.N. 1962*b* The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* **27**, 219–246.

Shepard, R.N. 1966 Metric structures in ordinal data. *J. math. Psychol.* **3**, 287–315.

Shepard, R.N. 1974 Representation of structure in similarity data: problems and prospects. *Psychometrika* **39**, 373–422.

Shepard, R.N. 1980 Multidimensional scaling, tree-fitting and clustering. *Science, Wash.* **210**, 390–398.

Sherman, C.R. 1972 Nonmetric multidimensional scaling: a Monte Carlo study of the basic parameters. *Psychometrika* **37**, 323–355.

Sibson, R. 1978 Studies in the robustness of multidimensional scaling: Procrustes statistics. *Jl R. statist. Soc.* B **40**, 234–238.

Sibson, R., Bowyer, A. & Osmond, C. 1981 Studies in the robustness of multidimensional scaling: Euclidean models and simulation studies. *J. statist. Computat. Simulat.* **13**, 273–296.

Simmen, M.W. 1994 Multidimensional scaling of binary dissimilarities: direct and derived approaches. *Multivar. Behav. Res.* (In the press.)

Simmen, M.W., Goodhill, G.J. & Willshaw, D.J. 1994 Scaling and brain connectivity. *Nature, Lond.* **369**, 448–450.

Sneath, P.H.A. & Sokal, R.R. 1973 *Numerical taxonomy*. San Francisco: Freeman.

Spence, I. & Ogilvie, J.C. 1973 A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivar. Behav. Res.* **8**, 511–517.

Stenson, H.H. & Knoll, R.L. 1969 Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychol. Bull.* **71**, 122–126.

Stevens, S.S. 1951 Mathematics, measurement and psychophysics. In *Handbook of experimental psychology* (ed. S. S. Stevens), pp. 1–49. New York: Wiley.

Takane, Y., Young, F.W. & de Leeuw, J. 1977 Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika* **42**, 7–67.

Torgerson, W.S. 1952 Multidimensional scaling, I: theory and method. *Psychometrika* **17**, 401–419.

Torgerson, W.S. 1986 Scaling and Psychometrika: spatial and alternative representations of similarity data. *Psychometrika* **51**, 57–63.

Ungerleider, L.G. & Mishkin, M. 1982 Two cortical visual systems. In *Analysis of visual behaviour* (ed. D. G. Ingle, M. A. Goodale & R. J. Q. Mansfield), pp. 549–586. Cambridge, Massachusetts: MIT Press.

Wagenaar, W.A. & Padmos, P. 1971 Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. *Br. J. math. statist. Psychol.* **24**, 101–110.

Wartenberg, D., Ferson, S. & Rohlf, F.J. 1987 Putting things in order: a critique of detrended correspondence analysis. *Am. Nat.* **129**, 434–448.

Wilkinson, E.M. 1971 Archaeological seriation and the travelling salesman problem. In *Mathematics in the archaeological and historical sciences* (ed. F. R. Hodson, D. G. Kendall & P. Tăutu), pp. 276–283. Edinburgh University Press.

Young, F.W. 1970 Nonmetric multidimensional scaling: recovery of metric information. *Psychometrika* **35**, 455–473.

Young, F.W. 1984 Scaling. *A. Rev. Psychol.* **35**, 55–81.

Young, F.W. 1987 *Multidimensional scaling: history, theory, and applications*. Hillsdale, New Jersey: Lawrence Erlbaum.

Young, F.W. & Harris, D.F. 1990 Multidimensional scaling: procedure ALSCAL. In *SPSS base system user's guide* (ed. M. Norusis), pp. 397–461. Chicago: SPSS.

Young, F.W. & Null, C.H. 1978 Multidimensional scaling of nominal data: the recovery of metric information with ALSCAL. *Psychometrika* **43**, 367–379.

Young, M.P. 1992 Objective analysis of the topological organization of the primate cortical visual system. *Nature, Lond.* **358**, 152–155.

Young, M.P. 1993 The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. Lond.* B **252**, 13–18.

Young, M.P., Scannell, J.W., Burns, G.A.P.C. & Blakemore, C. 1994 Scaling and brain connectivity – a reply. *Nature Lond.* **369**, 449–450.

Young, M.P. & Yamane, S. 1992 Sparse population coding of faces in the inferotemporal cortex. *Science, Wash.* **256**, 1327–1331.

Zeki, S. & Shipp, S. 1988 The functional logic of cortical connections. *Nature, Lond.* **335**, 311–317.

Received 17 June 1994; revised 1 September 1994; accepted 18 October 1994