PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Non-Metric Multidimensional Scaling in the Analysis of Neuroanatomical Connection Data and the Organization of the Primate Cortical Visual System

M. P. Young, J. W. Scannell, M. A. O'Neill, C. C. Hilgetag, G. Burns and C. Blakemore

Phil. Trans. R. Soc. Lond. B 1995 **348**, 281-308
doi: 10.1098/rstb.1995.0069

Non-metric multidimensional scaling in the analysis of neuroanatomical connection data and the organization of the primate cortical visual system

M. P. YOUNG*, J. W. SCANNELL, M. A. O'NEILL*, C. C. HILGETAG*, G. BURNS* AND C. BLAKEMORE

Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, U.K.

SUMMARY

Neuroanatomists have established that the various gross structures of the brain are divided into a large number of different processing regions and have catalogued a large number of connections between these regions. The connectional data derived from neuroanatomical studies are complex, and reliable conclusions about the organization of brain systems cannot be drawn from considering them without some supporting analysis. Recognition of this problem has recently led to the application of a variety of techniques to the analysis of connection data. One of the techniques that we previously employed, non-metric multidimensional scaling (NMDS), appears to have revealed important aspects of the organization of the central nervous system, such as the gross organization of the whole cortical network in two species. We present here a detailed treatment of methodological aspects of the application of NMDS to connection data. We first examine in detail the particular properties of neuroanatomical connection data. Second, we consider the details of NMDS and discuss the propriety of different possible NMDS approaches. Third, we present results of the analyses of connection data from the primate visual system, and discuss their interpretation. Fourth, we study independent analyses of the organization of the visual system, and examine the relation between the results of these analyses and those from NMDS. Fifth, we investigate quantitatively the performance of a number of data transformation and conditioning procedures, as well as tied and untied NMDS analysis of untransformed low-level data, to determine how well NMDS can recover known metric parameters from artificial data. We then re-analyse real connectivity data with the most successful methods at removing the effects of sparsity, to ensure that this aspect of data structure does not obscure others. Finally, we summarize the evidence on the connectional organization of the primate visual system, and discuss the reliability of NMDS analyses of neuroanatomical connection data.

1. INTRODUCTION

Structures in the central nervous system (CNS) typically make many connections with other structures. The primary visual cortex, V1, for example, exchanges substantial numbers of projection fibres with at least 60 other cortical and subcortical regions. This profuse connectivity is shared by most other structures in the CNS so that the brain, at one level, can be thought of as a complex wiring network. Considering only the cerebral cortical part of this network, there are approximately 1000 reported ipsilateral cortico-cortical connections areas in the cat and monkey (Scannell & Young 1993; Young 1993; Young *et al.* 1995). The complexity of even this part of the central nervous network is too great to allow conclusions about its organization to be drawn by unaided intuition. It has become an article of faith for many researchers, however, that connectivity data hold an important key to unravelling principles of brain organization (Rockland & Pandya 1979; Maunsell & Van Essen 1986; Livingston & Hubel 1988; Felleman & Van Essen 1991; Young 1992). The complexity of the data, and the promise that they hold for providing insight into the organization of central nervous processing, have encouraged the development and application of data analytic methods for treating connection data systematically. Several methods have been applied recently to the problem of untangling the CNS's connectivity, including hierarchical analysis (Rockland & Pandya 1979; Maunsell & Van Essen 1986; Felleman & Van Essen 1991), cluster analysis (Musil & Olson 1991), hodological analysis (Nicolelis *et al.* 1990), and non-metric multidimensional scaling (Shepard 1962, 1980; Young 1992, 1993; Scannell & Young 1993).

The latter approach using NMDS has been applied more widely than any other method. Every major sensory system in the cat and monkey cortex, and the

* Present address: Neural Systems Group, Department of Psychology, Ridley Building, Newcastle upon Tyne, NE1 7RU, U.K.

gross cortical organization of each animal, have been analysed in this way (Young 1992, 1993; Scannell & Young 1993; Young *et al.* 1995). In this approach it is considered that connection data indicate the proximities beween brain structures in a notional high-dimensional space, and NMDS is used to reduce the number of dimensions of the space to make the configuration defined by the connections more understandable. The results of these analyses indicate that all the major cortical sensory systems in cats and monkeys are hierarchical, serially ordered structures. In addition, the NMDS analyses of all the available cortical connection data seem to show that both the monkey and cat cortical networks are divided into three major hierarchically organized sensory systems and a fourth system composed of prefrontal and limbic structures, which is connectionally distant from the sensory–motor periphery (Young 1992, 1993; Scannell & Young 1993; Young *et al.* 1995). The analyses also appear to show that the primate visual system is clearly divided into two gross streams, unlike any other central sensory system (Young 1992).

These findings are potentially significant and it is important, therefore, to know whether the results of the analyses are reliable. We have not previously treated methodological aspects of the application of NMDS to neuroanatomical connection data in detail. We do so here by:

1. Examining the particular properties of neuroanatomical connection data, and determining the types of underlying structure that are consistent with the data's simple statistics.
2. Considering the details of NMDS and discussing the propriety of different possible NMDS approaches to data with these particular properties.
3. Presenting detailed results of the analyses of connection data from the primate visual system, and discussing their interpretation.
4. Turning to independent analyses of the organization of the visual system, and examining the relation between the results of these analyses and those from NMDS.
5. Investigating quantitatively the performance of a number of data transformation and conditioning procedures, as well as tied and untied NMDS analysis of untransformed low-level data, to determine how well NMDS can recover known metric parameters from artificial data. We then re-analyse real connectivity data with the most successful methods to ensure that the data's sparsity has not obscured important aspects of data structure.
6. Summarizing all the evidence on the connectional organization of the primate visual system, and discussing these facts in the context of the conclusions derived from application of NMDS to the primate visual system data (Young 1992).

2. CHARACTERISTICS OF THE DATA

Any approach to data analysis properly begins by examining the characteristics of the data that are to be analysed. So what are the characteristics of connection data?

Many different neuroanatomical methods have been employed to reveal connections between brain structures. These methods have included the dissection of cerebral white matter (Vieussens 1685; Dejerine 1895), strychnine neuronography (Pribram & MacLean 1953), and the Nauta–Gygax (Whitlock & Nauta 1956) and Fink–Heimer methods (Turner *et al.* 1980) of tracing neuronal degeneration after lesion. More recently, the active transport of tracer chemicals by neurons from a point of injection either anterogradely (to the synaptic targets of the neurons whose cell bodies lie at the injection site), or retrogradely (to the cell bodies of neurons whose axonal processes terminate in the injection site) has become the method of choice. Actively transported methods include autoradiography, often using tritiated amino acids (Amaral & Price 1984; Ungerleider & Desimone 1986), horseradish peroxidase (Aggleton *et al.* 1980), carbocyanine dyes (Vidalsanz *et al.* 1988), other fluorescent dyes (De Yoe & Van Essen 1985), Phaseolus leucoagglutinin (Gerfen & Sawchenko 1984), biocytin (King *et al.* 1989), and fluorescent microbeads (Katz & Iarovici 1990).

All the modern neuroanatomical methods that yield results bearing on the connections between brain structures involve sectioning or flattening the brain and visualizing the distribution of transported label. These methods thus identify the connections, and the direction of connections, between different brain regions. The brains of a number of mammals have been extensively investigated by these means. Two of the best studied mammals, the cat and the Macaque monkey, have cerebral cortices that contain about 70 distinct areas linked by about 1000 reported connections (Scannell & Young 1993; Young 1993). Many of these areas possess marked internal structure, in which their subcompartments have distinct patterns of connectivity (Felleman & Van Essen 1991), and many areas lying in what used to be termed 'association cortex' are not clearly demarkated from their neighbours (Colby & DuHamel 1991). Some structures are more widely connected than others (the amygdala in the macaque, for example) but the 'average' cortical area sends or receives nerve fibres from about 25% of the ipsilateral cortical areas, and does not send or receive an appreciable number of fibres from the remaining 75% of areas (Felleman & Van Essen 1991; Scannell & Young 1993; Young 1993). The strength or density of the connections varies greatly. Large robust connections (such as the projection from primary to secondary visual areas) may consist of many tens of millions of fibres whereas the weaker, but still identified, connections may contain many times fewer fibres (Olson & Musil 1992). Areas that are considered by neuroanatomists not to be connected may exchange a few fibres, sometimes revealed as 'background labelling', but the number of fibres exchanged is typically many orders of magnitude lower than for interactions reported as connections.

The ideal type of connection data for analysing the area-to-area pattern of connectivity would concern the number of nerve fibres linking any two areas. This kind of data would perfectly capture the relations between

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

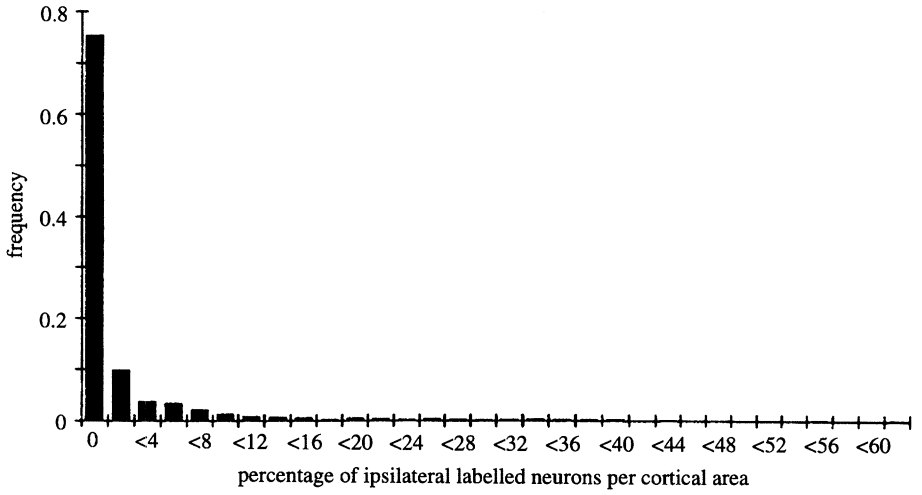

Figure 1. Quantitative distribution of neuroanatomical connection strengths. This distribution was derived from data from 26 injections into areas of cortex and was compiled from five quantitative neuroanatomical studies (see text). Cases in which there is no significant tracer transport are the most frequent, a result that reflects the sparsity of connection at all levels of the central nervous system. The quantitative distribution of connection strength that is approximated here provides important constraints on the possible structures that can be implemented in cortical systems.

brain regions that are defined by their patterns of connectivity, and would provide a metric proximity measure. It is possible to estimate the absolute number of nerve cells in the optic nerve (10^6), or sending projections from V1 to V2 (25×10^6; Van Essen & De Yoe 1995), but there are not enough data to reveal the number of nerve cells involved for the vast majority of connections that are known to exist. However, within a particular anatomical study it is sometimes possible to estimate the relative numbers of connections passing, for example, into one area in which an injection has been made from several other areas. These data often take the form of the percentage of the total number of labelled cells found in each area.

(*a*) *The distribution of projections of different strength*

Figure 1 shows the distribution of neuroanatomical connection strengths derived from 26 injections into six areas of cortex. The data were taken from five studies that presented information on the percentage of labelled neurons in different areas following injection of the retrograde tracers Nuclear Yellow and Bisbenzimide (Olson & Jeffers 1987; Olson & Lawler 1987; Bowman & Olson 1988; Musil & Olson 1988*a*,*b*; Olson & Musil 1992). The studies systematically scanned sections for labelled neurons and expressed the number of neurons found in each area as a percentage of the total labelled neurons in the hemisphere ipsilateral to the injection site and outside the injected cortical area. The number of neurons counted per hemisphere ranged from 1000 to 10000, and the percentage of labelled neurons per area ranged from 60–0%.

Figure 1 represents an estimate of the average frequency distribution of retrogradely labelled neurons after injection in a cortical area, and may be interpreted as an estimate of the distribution of relative connectional proximity between cortical areas. It is clearly apparent that the most frequent connectional proximity is zero. This reflects the fact that the cortex is remarkably sparsely connected. There are about 10^{10} neurons in a mammalian cortex, each of which have about 1000 connections with other neurons. These values mean that, at the neuronal level, the probability that any two randomly selected neurons are connected is about one hundred thousandth of one percent. Figure 1 represents the aggregate level of connections between brain areas and shows that, even at this level, connectivity is sparse with the majority of possible connections (often around 70%) being absent. Figure 1 also illustrates that strong connections are few in number, and may lie in a region of the distribution that is descending asymptotically to a frequency of zero. The relative frequency of different connectional proximities provides a clue to the structure or shape that reflects the cortical areas' underlying geometry. Figure 1 indicates that the connectional structure contains a high proportion of long distances (zero connectional proximities) with few medium or short distances. This is important information, as it is diagnostic of a family of shapes to which the underlying configurations of neural systems belong (see §2*c*).

(*b*) *Connection matrices*

Analysis of the area-to-area pattern of connectivity in the cortex requires a connection matrix that summarizes the connections between an interestingly large set of brain regions. Such a matrix must be constructed from many separate reports of connectivity from many different laboratories using many different anatomical methods. Individual anatomical studies are seldom sufficient. It cannot reasonably be assumed that the estimates of the relative proportions of labelled cells, discussed above, hold across different studies. There are in any case, too few quantitative studies of this nature to define a proximity matrix at this high level of measurement.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

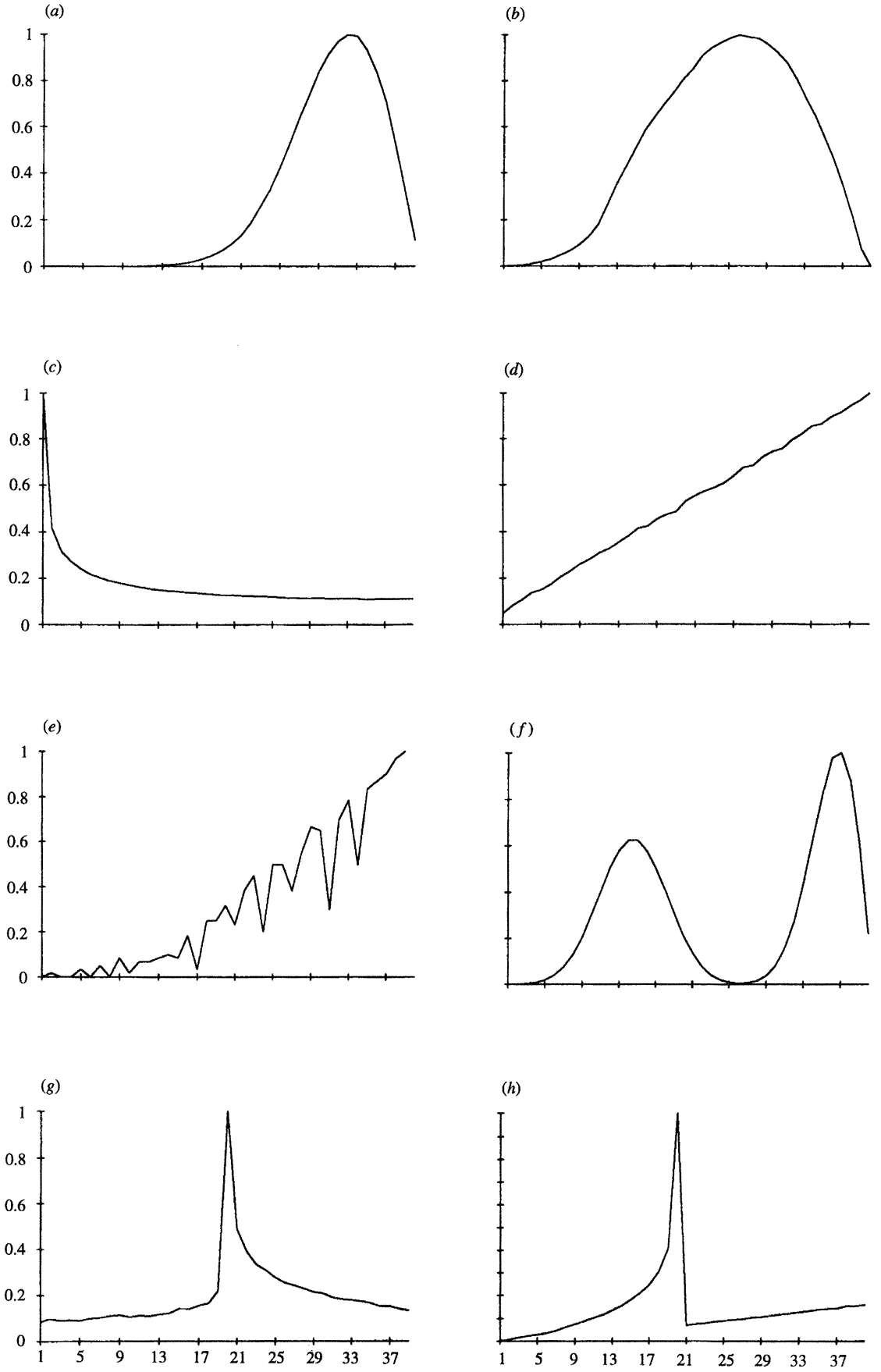

Figure 2. Distributions of distances that are associated with different shapes. The figure shows the frequency of interpoint distances between the points of eight example Euclidean structures. The *x*-axis corresponds to distances between points in the arbitrarily sized structures. The structures are (*a*) a Gaussian patch, (*b*) a disk region within

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Quantitative connection data are not therefore sufficiently widely available to define an interval connection matrix (see Coombs (1964) for an exposition of an influential theory of measurement). There is, however, a good deal of information in the neuroanatomical literature concerning whether particular projections are strong, moderate or sparse, particularly in the cat. These classifications of connection density are assumed to correspond to neuroanatomists' judicious divisions of the distributions of the density of labelled cells that they see in their sections. If it is assumed that the differences between different studies are not so gross as to cause one study's strong connections to be equivalent to another study's sparse ones, then it is possible to summarize the current state of knowledge of patterns of cortical connectivity in connection matrices with an ordinal ranking of connection strength. Scannell & Young (1993) used this approach for cat cortical connectivity and ranked the connections as 3 (strong), 2 (intermediate), 1 (sparse) or 0 (absent or unreported). When reflected about the leading diagonal of the connection matrix, this provided a 7 point rating scale (0–6) describing the connectional proximity of areas of the cat brain.

Information on the strength of connection between structures is not available for every reported connection, however. In many cases, but particularly in the case of the monkey, only the fact of connection (or the fact of the lack of a connection) is reported. Young (1992, 1993) summarized the reported data on monkey cortical connectivity by giving interactions between areas a rank of 2 if the areas exchanged reciprocal connections, 1 if the areas were connected in only one direction, or 0 if connections between the areas were reported absent or no connections were reported, giving a 3 point rating scale for connectional proximity.

Non-zero entries in either the 3 point or 7 point connection matrices correspond to a range of 'real' metric proximities. This is because the non-zero entries are estimates of the rank order of the strength of connection between areas, which in turn depend on the number of nerve fibres linking any two areas. The number of fibres passing between areas will differ, even when two connections are assigned the same value in the connection rating scale. There is, however, no basis for estimating these finer differences in any analytical procedure. The only reliable means of discriminating these small distinctions in proximity is to undertake further empirical neuroanatomical work. For this reason, we consider that no attempt should be made to assign different values to entries with the same rank. This has important consequences for the appropriate analytical strategy (see §3*b*).

In contrast to the reported connections, there is no supportable sense in which the zero entries correspond to a range of metric proximities. The zero entries do not correspond to a range of real metric proximities because the non-existence of a connection is not a graded quality. It is meaningless to say of one absent connection that it is more or less absent than another absent connection. This is another way of saying that should there be any differences between the background labelling observed for two reported non-connections, then these differences should not be respected. The number of fibres exchanged in such background labelling cases is many orders of magnitude lower than for *bona fide* connections, and any differences in background label are not considered meaningful by the neuroanatomists who report them. Zero entries corresponding to possible connections that have been confirmed absent in anatomical studies therefore signal the same low proximity between the areas to which they relate.

Some of the zeros in connection matrices correspond to possible connections whose existence or absence has not been explicitly reported. Should this subset of the zero entries be treated in a different way to those that correspond to possible connections that have been confirmed absent? It might be tempting to perform a kind of missing data estimation on these entries, perhaps by assigning to particular examples of these entries a probability that the particular possible connection exists. We note, however, that this approach would entail second-guessing neuroanatomy and neuroanatomists. Connectional neuroanatomy is not predictable. We therefore regard any such missing data estimation procedure as dangerously likely to mislead. We consider that a better approach is to analyse a 'control' data matrix, constructed under the assumption that unreported connections all exist (the grossest possible perturbation of the data), to see the extent to which the uncertainties in the connection data could give rise to different conclusions as more data are reported.

Connection data thus present themselves as sparse similarity matrices, in which the non-zero entries specify that the particular connection is estimated to involve a particular range of connection density and in which the zero entries indicate that all areas not exchanging connections are exactly as far apart as each other. These characteristics of connection data, some of which are unusual, have important consequences for choosing appropriate data analytic methods (see §3*b*), and for interpreting NMDS configurations derived from neuroanatomical connection data. Before proceeding to methods of data analysis, we return to the relative frequency of different connectional proximities from quantitative neuroanatomical studies, and consider further the consequences of this data distribution.

(*c*) *The underlying connectional 'shapes' of neural systems*

Figure 1 indicated that the connectional data contain a high proportion of zero connectional proximities, with few medium or high proximities. This

which points are uniformly but randomly distributed (*c*) a ring, (*d*) a straight line, (*e*) a regular hexagonal grid, (*f*) three Gaussian patches whose centres are placed at the three vertices of an equilateral triangle, (*g*) two rings in a figure-of-eight configuration and (*h*) parallel straight lines.

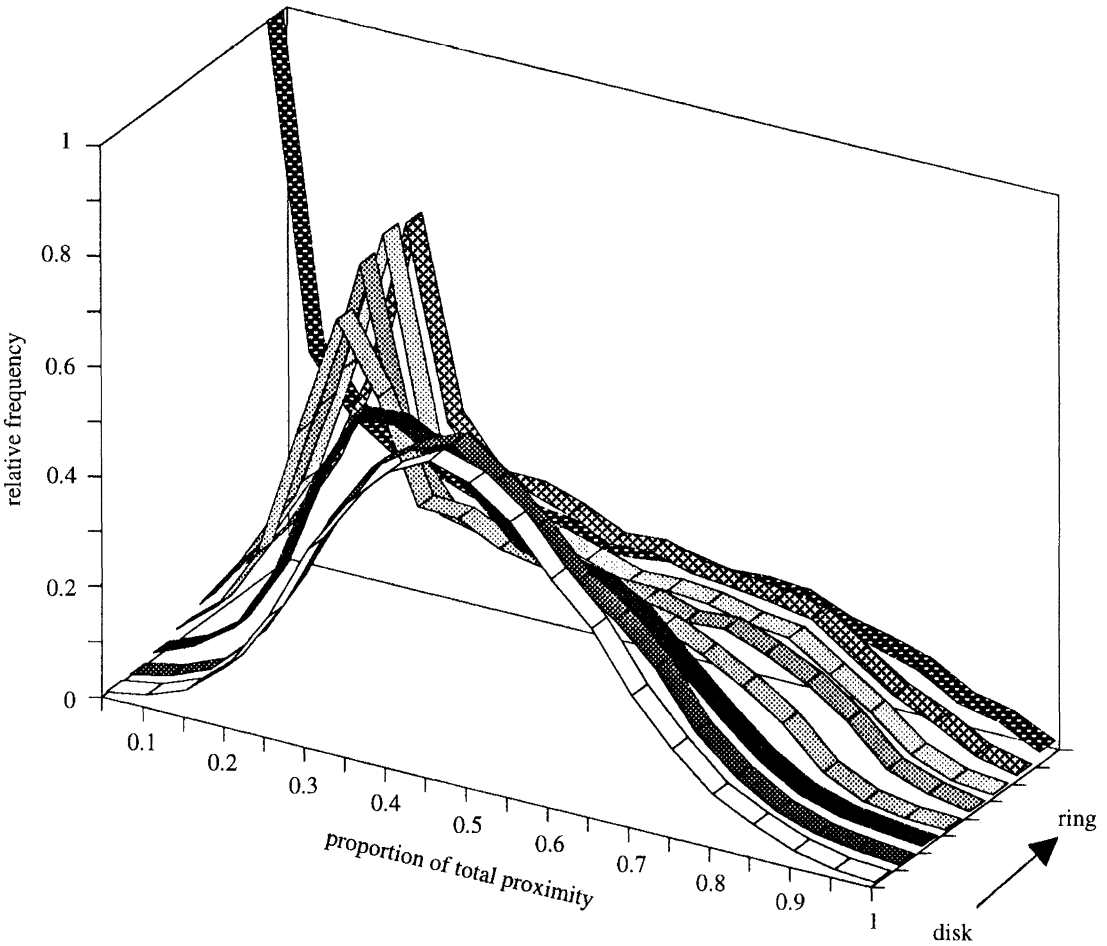

Figure 3. Distributions of relative proximities for points lying randomly within a disk, and in successively more ring-like structures, as an empty region is gradually expanded in the centre of the space. In the limit case, depicted as the rearmost of the distributions, where the points are distributed on the circumference of a circle, the distribution of relative proximity between points is the first differential of the inverse of the cosine of an angle between the circle's diameter and a chord passing between 0° and 90 °. $[-(1 - x^2)^{-1/2}]$. All distributions are shown as transformed by subtracting from all distances the maximum distance present in the structure, and then expressing each proximity value as a fraction of all the proximities present in the row.

information provides important constraints on the family of shapes which the underlying configurations of neural systems can occupy. Consider that if points are arranged in a space and the distances between them measured, the distribution of distances between the points depends on the shape of the arrangement of points. Figure 2 illustrates some of the different distributions of distances that are associated with different shapes. The figure shows the relative frequency of interpoint proximities between the points of eight example Euclidean structures. The structures are: (i) a Gaussian patch;(ii) a disk region within which points are uniformly but randomly distributed; (iii) a ring; (iv) a straight line; (v) a regular hexagonal grid; (vi) three Gaussian patches whose centres are placed at the three vertices of an equilateral triangle; (vii) two rings in a figure-of-eight configuration; and (viii) parallel straight lines. The disk, Gaussian, and line contain relatively high frequencies of medium or high proximities between points. The ring contains relatively high frequencies of low proximities. The three Gaussian patches shape shows a bimodal distribution of distances, corresponding to inter- and intra-patch distances. The two rings shape shows a sharp peak (at the distance corresponding to the diameter of the rings), where proximities greater than this peak correspond to intra-ring distances, and proximities lower than this peak correspond to inter-ring distances. The parallel lines shape also shows a sharp peak, whose position corresponds to the line spacing. The regular hexagonal grid shows several peaks at medium distances. The positions of the peaks correspond to harmonics of the internode distance in the grid. It is apparent that the frequency distributions of distances or proximities can alone give an important indication of the underlying organization of the data structure. Any structure, for example, with a large proportion of long distances or low proximities cannot be discoid, Gaussian, figure-of-eight, or involve extended straight lines or regular grid-like organization.

It is instructive to compare these test distributions and shapes directly with the distribution for neuroanatomical connection data presented in figure 1. It is necessary first to perform a simple transform on the test data to make this direct comparison. The quantitative neuroanatomical connection strengths in figure 1 were given as a percentage of labelled cells per hemisphere. This corresponds to the fraction of the total proximity

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

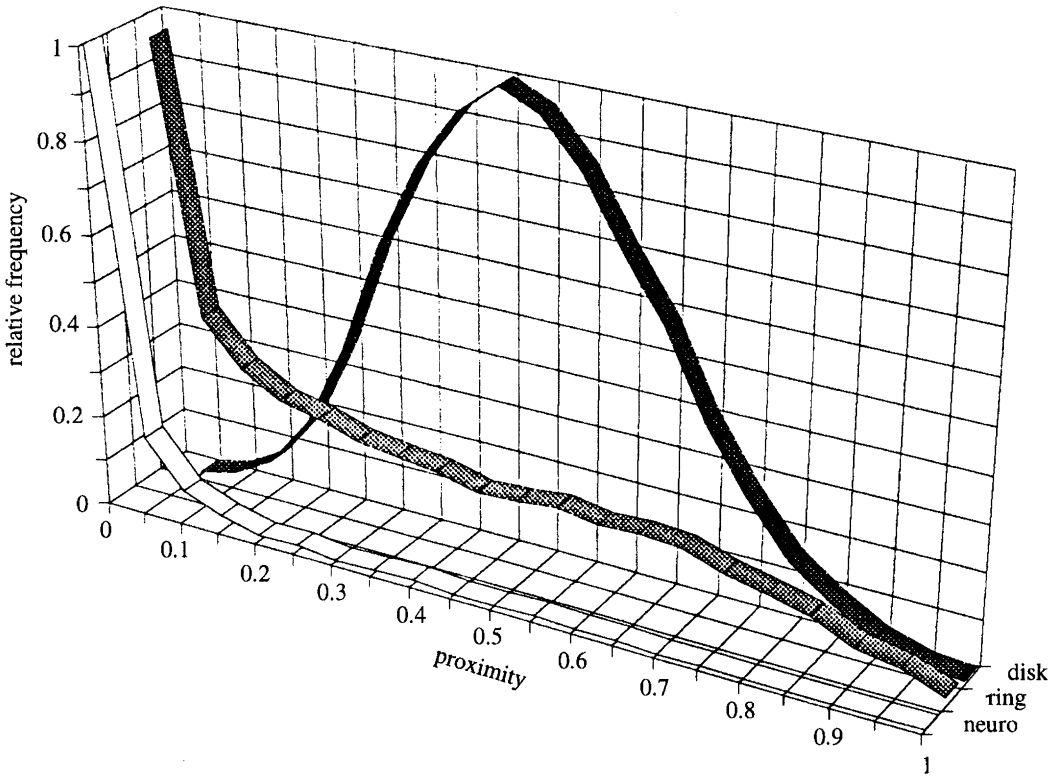

Figure 4. Distributions of relative proximity for disk, ring and quantitative neuroanatomical data. The neuroanatomical distribution is well approximated by the ring distribution, but bears no relation to the distribution for disk data. These correspondences are given quantitative statistical expression in the text.

that each connection represents: if the connection to one area contains 60%, no other area can contain more than 40% of the labelled neurons. For direct comparison, it is necessary to convert the metric distances between the points of the test structures into the same form. We therefore converted the metric distance matrices into proximity matrices by subtracting from all distances the maximum distance present in the structure, and then expressing each proximity value as a fraction of all the proximities present in the row. We paid particular attention to ring and disk test structures because both have been suggested as the underlying connectional shape of particular neural systems (Young 1992; Goodhill *et al.* 1994).

Figure 3 summarizes the distributions of relative proximities for points lying randomly within a disk, and in successively more ring-like structures as an empty region is gradually expanded in the centre of the space. Structures in which the points are distributed uniformly within a circular region possess a unimodal distribution of relative proximity whose peak is at middle proximities. As the proportion of the central region from which points are abolished increases, i.e. as the configuration becomes an increasingly thin ring structure the structure comes to contain more low, fewer high, and many fewer medium proximities. In the limit case, where the points are distributed on the circumference of a circle, the distribution of relative proximity between points is the first differential of the inverse of the cosine of an angle between the circle's diameter and a chord passing between 0–90°, transformed in the manner described above.

The distribution of neuroanatomical connection strength is replotted in figure 4 alongside representative distributions from disk and ring structures to aid direct comparison. The distribution of proximities in the quantitative neuroanatomical data is diagnostic of a ring or horseshoe structure. The high proportion of low proximities in the neuroanatomical connection data compels the conclusion that the underlying shape of neural systems cannot be that of a uniform disk (cf. Goodhill *et al.* 1994). These relations between the proximities in anatomical data and those of the test structures can be rendered in quantitative terms by regressing the one on the others. The relation between the connection data's distribution and that of the disk was not statistically significant ($r^2 = 0.09$, $p = 0.2$), while the relation with that of the ring was statistically significant ($r^2 = 0.88$, $p = 0.0001$).

Thus, a consideration of the quantitative characteristics of anatomical connection data, before any consideration of detailed methods of data analysis, indicates that the Euclidean configuration of points that closely approximates the connection data for any cortical neural system has an annular or horseshoe shape. Similar considerations apply in the case of dimensions greater than two, as we shall report elsewhere. This conclusion does not bear on the issue of where particular brain structures should lie in this manifold. It therefore says nothing about whether connection data imply a structure with a number of streams or only one, or about the ordering of brain areas relative to one another. To address these questions, the patterns of connectivity embodied in the connectional constraints themselves require to be

optimized, and so we turn to NMDS, a procedure that performs precisely this operation.

3. NON-METRIC MULTIDIMENSIONAL SCALING

NMDS is a method of data analysis that constructs a spatial representation of a set of elements on the basis of a table of 'proximities' that define the relations between the elements. Historically, NMDS is related to a number of earlier statistical procedures, often referred to as principal coordinates analysis or metric multidimensional scaling (Torgerson 1952). These procedures were designed to take metric distance measures and to produce their corresponding metric coordinates. Metric methods of this kind are limited in application due to the assumption that the proximity data to be analysed are metric in character (Shepard 1980). This restriction invited the development of several varieties of non-metric multidimensional scaling, in which the intention was to overcome the problems of the metric approach either by estimating the form of the function that mapped the proximities in the input onto the Euclidean distances in the output (Shepard 1958), or by restricting the goal of the analysis to finding the ordering of the elements on orthogonal axes of a space (Coombs 1964).

The NMDS algorithms that are now widely used arose from a distinctive approach based on 'analysis of proximities' (Shepard 1980). Here, an iterative procedure is used to adjust the positions of points in the output space until the ordering of the distances between the points is as close as possible to the reverse ordering of the corresponding proximities (Shepard 1962). The sense in which distances and proximities were 'as close as possible' found an explicit form in a 'sum of squares' measure of departure from a perfect monotonic relation between them called STRESS (Kruskal 1964*b*). This development brought the approach to NMDS arising from analysis of proximities almost to its current state of development by allowing the optimization of non-metric constraints by standard gradient descent methods.

(*a*) *Multivariate descriptions of a set of elements from the proximities between them*

NMDS pursues the following procedure to find a metric configuration of n points that optimally represents the non-metric constraints in an $n \times n$ input matrix. The procedure begins by placing an initial configuration of n points in a space with K specified dimensions, where K is an interpretably low number. This placement may be made either at random, or by the application of principal coordinates analysis. A set of numbers, termed disparities, δ_{ij}, are defined that enjoy a monotonic relation with the input proximities s_{ij}, while also fitting the distances in the configuration d_{ij}. The program then iterates toward an optimal stationary configuration.

Each iteration involves three elements. First, the best fitting sequence of disparities δ_{ij} is determined by monotone regression (Kruskal 1964*b*). Second, the $n \times K$ partial derivatives of STRESS (s) implied by the current coordinates of the points are evaluated by

$$S = \left[\sum_{i,j} (d_{ij} - \delta_{ij})^2\right] \Big/ \sum_{i,j} d_{ij}^2,$$

where the distances δ_{ij} are normally Euclidean distances in K dimensions, given by

$$d^{ij} = \left[\sum_{k=1}^{k} (x_{ik} - x_{jk})^2\right]^{\frac{1}{2}},$$

although other distance measures, corresponding to non-Euclidean spaces, can be chosen (Kruskal 1964*b*).

Third, the current coordinates are moved in the direction of negative gradient, calculated by

$$x_{ik}^{\text{new}} = x_{ik}^{\text{old}} - \alpha(\delta S / \delta x_{ik})$$

in which α is a step-size factor (Shepard 1980). The iterations of the procedure terminate either when the STRESS gradient is sufficiently small as to indicate that the convergence on a solution is near stationary, or when a criterion number of iterations have been performed, and the resulting parameters are written as output.

More recently, a very efficient method for performing NMDS has been described (de Leeuw *et al.* 1976), which its originators call ALSCAL. This method uses 'alternating least squares', an algorithm that is necessarily convergent, quick, and relatively free from the problems of stalling in local minima (Takane *et al.* 1977). We have preferred to use this algorithm, but for computational efficiency the program uses a different objective cost function, called SSTRESS:

$$SS = \left[\sum_{i,j} (d_{ij}^2 - \delta_{ij}^2)^2 \Big/ \sum_{i,j} d_{ij}^4\right]^{\frac{1}{2}}.$$

It is important to note that large disparities, that is, those related to the zero proximities in the case of connection data, are emphasized over smaller disparities by both STRESS and SSTRESS. This is because the costs of misatches between distances and disparities are calculated in equations that include them as a quadratic term for STRESS, and as a quartic term for SSTRESS. Other procedures that emphasize local monotonicity rather than global monotonicity have been described (e.g. PARAMAP, Shepard & Carroll 1969), but the algorithms that implement these methods are not yet capable of dealing reliably with data at the level of measurement represented by connection data.

A further aspect of the NMDS procedure concerns the way the algorithm deals with similarities that possess the same value. There are two alternatives in this circumstance, namely the primary or untied approach, and the secondary or tied approach. The difference between these two approaches lies in the different constraints under which the disparities δ_{ij} corresponding to tied similarities are placed. In the primary, untied, approach, when the s_{ij} and s_{kl} are equal, their corresponding disparities δ_{ij} and δ_{kl} are not constrained beyond requiring that whenever $s_{ij} < s_{kl}$, $\delta_{ij} \leqslant \delta_{kl}$. In this case, the terms $(d_{ij} - \delta_{ij})^2$ and $(d_{kl} - \delta_{kl})^2$ are

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

allowed to be zero (not to contribute to the cost), unless they are prevented from being so by other constraints. Hence non-equivalence of disparities that correspond to the same similarity value is not penalized in the optimization procedure, and the distances in the output configuration that correspond to tied similarities are allowed to vary between limits prescribed by the other similarity levels (Kruskal 1964*b*). This may be summarized by

Whenever $s_{ij} < s_{kl}$, then $\delta_{ij} \leqq \delta_{kl}$.

In the secondary, tied, approach, when the disparities s_{ij} and s_{kl} are equal, their corresponding distances should also be equal. In this case, the terms $(d^{ij} - \delta_{ij})^2$ and $(d_{kl} - \delta_{kl})^2$ are not allowed to be zero (and so contribute to the cost function if d^{ij} and d^{kl} are unequal). Hence, distances in the output configuration that correspond to tied similarities are penalized when they are not also equivalent (Kruskal 1964*b*). This can be summarized by the constraints

Whenever $s_{ij} < s_{kl}$, then $\delta_{ji} \leqq \delta_{kl}$.

Whenever $s_{ij} = s_{kl}$, then $\delta_{ij} = \delta_{kl}$.

The choice of whether to follow the tied or untied approach is often governed by the question of whether the distribution that underlies the data is continuous or discrete (Takane *et al.* 1977). This choice is one of those that must be made in selecting the appropriate NMDS approach to neuroanatomical data.

(*b*) *What is the appropriate NMDS approach to neuroanatomical connection data?*

We described aspects of the characteristics of the connection data in §2. The task of this section is to decide the appropriate parameters for NMDS analysis of data with these properties. Because connection data matrices always contain many entries with the same value, the first decision concerns whether the tied or untied approach in NMDS should be taken.

Quantitative neuroanatomical studies, such as those that informed the discussion in §2*a*, suggest that connection strengths are distributed in a continuous manner (see figure 1). The rating scales in connection matrices, however, represent this smooth dispersion by discrete values. The continuous distribution of connection density is cut into discrete regions, each signalled by a discrete value, by placing thresholds. To capture accurately the underlying structure of the data these cusps between regions of the distribution must be placed intelligently, and NMDS must then reconstruct the proximities to which the discrete values correspond. The first of these requirements will be met if neuroanatomists' classifications of the labelling in their sections are judicious (see §2*b*). The second will be met if NMDS faithfully reconstructs the proximities signalled by the discrete values in the connection matrix.

These refections might imply that the primary, untied, approach should be taken in NMDS analysis of connection data. Two overriding considerations, however, council against this choice. First, as discussed in §2*b*, while the non-zero values in connection matrices certainly signify a range of proximities related to the number of nerve fibres linking any two areas, there is no basis for estimating these finer differences in a subsequent analysis. Resolution of these small distinctions in proximity requires further empirical neuroanatomical work. The analysis should consequently be limited to reconstructing the parameters of the data as presently given, rather than performing what would amount to a missing data estimation. For this reason, we consider that attempts to assign different values to entries with the same rank should be treated very cautiously, and therefore that the tied approach is the more appropriate.

Second, we indicated in §3*a* that NMDS lays emphasis on the large disparities carried by low proximities in the connection matrix. This means that the zero entries are of particular salience to the stationary configuration which will be found by NMDS. There is no interpretable sense in which the zero entries correspond to a range of proximities lying in some continuous distribution. Once classified by neuroanatomists as absent, it is not meaningful to say of two absent connections that the one is less absent than the other. The zero entries are therefore fundamentally unlike the non-zero entries, in that the part of the distribution of connection strength from which they derive does not possess an underlying continuum. Those zeros in connection matrices that correspond to possible connections whose absence has not been explicitly reported, however, could be considered to be drawn from a continuous distribution, where the continuity lies in different probabilities that particular possible connections exist. With reference to the above, an attempt to assign different values to unreported connections would be to perform a kind of missing data estimation on these entries, when there is no basis for such estimation. Connectional neuroanatomy is not predictable, and any such missing data estimation procedure is likely to be misleading. The emphasis that NMDS lays on the zero proximities, and the unusual property that all zero proximities should be reflected in distances of the same length, suggest that the tied approach is the appropriate one for these data.

Another choice concerns the geometry of the output space. Choosing non-Euclidean output spaces has been suggested to us as a means of mitigating the solutions' emphasis on larger disparities. We feel, however, that only Euclidean solution spaces will be readily interpretable by the human visual system, which is adapted by experience to a Euclidean three-space. We therefore consider Euclidean output spaces to be the most appropriate, on the grounds of interpretability (see Shepard 1980).

4. APPLICATION OF NMDS TO THE PRIMATE VISUAL SYSTEM

We have examined a matrix of connections between areas of the primate cortical visual system (Young 1992). The constitution of the matrix was derived from a collation of data presented by Felleman & Van Essen (1991), with a number of minor changes. We did not include areas MIP and MDP in the matrix because of lack of connection data, and because of some un-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

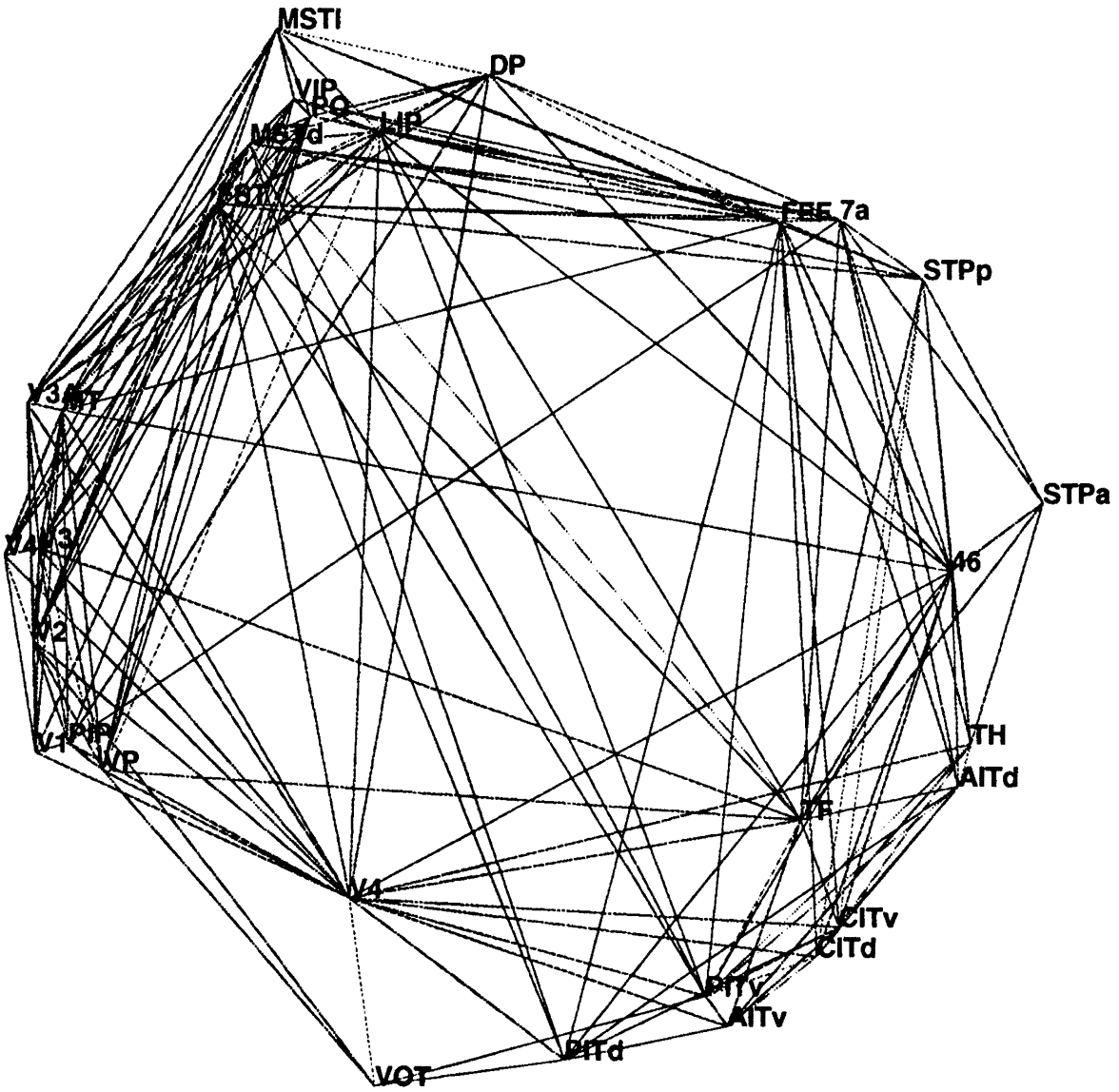

Figure 5. The structure derived from tied NMDS analysis of the visual system matrix. The features of this structure are described in the text, and in Young (1992). Briefly, the structure implies that the visual system is divided into two gross streams (both of which are hierarchically organized) and which reconverge in areas of the temporal and frontal lobes.

certainty about the relationship between these areas and area 7m of Cavada & Goldman-Rakic (1989). We assigned those connections of the posterior inferotemporal cortex (PIT) whose origins were not differentiated into the ventral and dorsal parts of PIT to both these subregions. This assigned connections with MSTD, FST, FEF and area 46 to both PITV and PITD. A second set of assignments was made for connections listed as related to CIT. This assigned connections with STPp, TH, FEF and area 46 to both CITV and CITD. A third set of assignments was made for connections listed as related to STP, so that STPp and STPa acquired connections with area 7a. Finally, the sparse connection between MT and area 46 (Barbas 1988) is credited with existence, despite its not having been reported by Ungerleider & Desimone (1986). The resulting matrix was reflected about its leading diagonal to derive a lower-triangular matrix with entries taking the values 2 (reciprocal), 1 (unidirectional), and 0 (non-existent or unreported) (see §2*b*), and has been presented in Young (1992). Because the task of this article is to address methodological aspects of the application of NMDS to connection data, we consider here exactly the same matrix of connections, rather than reporting analysis of matrices that are larger, or which contain more recent anatomical data.

After our discussion of the appropriate parameters for NMDS analysis of neuroanatomical connection data in §3*b*, we performed the analysis with tied similarities remaining tied, and with a Euclidean output space. The analysis was performed by program ALSCAL (de Leeuw *et al.* 1976; Takane *et al.* 1977). We review the characteristics of the resulting configuration (see Young 1992 and figure 5).

The points of the configuration are concentrated into an annular region of the output space, as expected from consideration of the quantitative aspects of connection data, in §2*a*,*c*. The annularity of the solution follows from the large proportion of non-connections in

the connection matrix and the underlying quantitative distribution of connection strength, and it is encouraging that the shape of the solution is among the candidate manifolds for which the quantitative distribution of neuroanatomical connection strength is diagnostic (§2*c*). Had the shape of the solution been that of a uniformly filled disk, for example, we would have cause for concern, because such a configuration would not be among the possible structures permitted by the quantitative distribution of connection strength.

The dimensions of the solution correspond approximately to the anterior–posterior and to the dorso–ventral (top to bottom) distribution of the areas spatially within the brain. Parietal cortical areas, for example, are placed toward the top of the diagram, whereas inferotemporal cortical areas are placed toward the bottom. If we remember that information on area-to-area connectivity was the only type of information that entered the analysis – no information regarding the disposition of the areas on the cortical sheet was included – this aspect of the configuration implies that the spatial location of an area well predicts the areas to which that area is likely to be connected and therefore, that nearby areas tend to exchange connections with one another. This offers some support to the idea that a constraint on the organization of the brain may be the requirement that neuronal wiring be kept to a minimum (Cowey 1979; Mitchison 1991; Young 1992; Cherniak 1994).

Primary visual cortex (V1) is located at the far left of figure 5. Signals are relayed from V1 to a group of prestriate areas including V2, V3, VP, V4t, V3A, MT, and, surprisingly because it is a posterior parietal area, PIP. MT and V3A are placed further from the sensory periphery than other members of this group. MT is further distinguished from its topological neighbours by its (sparse) projection to frontal cortex area 46 and its (less sparse) projection to the frontal eye fields (FEF) (see Felleman & Van Essen 1991). Every area in the prestriate group projects to a cluster of areas consisting of areas of the posterior parietal cortex and the caudal superior temporal sulcus, namely FST, MSTD, MSTL, VIP, PO, LIP and DP. These areas then project to FEF, parietal area 7a, the posterior region of the superior temporal polysensory area (STPp), and to area 46 and anterior STP (STPa).

Beginning at V1 again, but now concentrating on the lower part of the configuration, V1 projects to V4, while V2 and VP project to VOT. Signals are relayed from V4 and VOT into the areas of the inferotemporal (IT) cortex. The IT areas appear to be serially organized, with more anterior areas generally being placed successively further toward the right of the diagram, away from the sensory periphery. The higher-order areas of the IT cortex are associated with areas TF and TH of the parahippocampal cortex. The topologically higher-order IT areas project to STPa and to area 46.

It is a feature of the structure that relatively few connections pass across the central region between the parietal and inferotemporal groupings of areas, by comparison to the number that pass around the rim. Hence, there appear to be two distinct sets of areas in the cortical visual system that are much more profusely interconnected within groupings than between them (see §5*a*). These two sets of areas correspond straightforwardly to the dorsal and ventral streams of processing, which were proposed most clearly by Ungerleider & Mishkin (1982) on the basis of the behavioural effects of cortical lesions (see §5*d*). The higher-order areas of both streams project to STPa and to area 46. This feature of the structure implies that there is reconvergence of processed visual information in the rostral parts of the temporal lobe and in the frontal lobe (see Perrett & Oram 1995).

These aspects of the results of NMDS analysis of the primate cortical visual system indicate that, at the aggregate level of connections between gross brain areas, four principles underlie its organization: (i) neighbouring areas tend to exchange connections; (ii) it is dichotomized into two streams; (iii) both streams are broadly hierarchical; and (iv) the streams reconverge in area 46 and STPa (Young 1992).

(*a*) *Control results*

In this section we report details of a variety of further analyses to assess the reliability of the above conclusions. The first 'control' analysis concerns the possible effects of changing the status of those zero entries in the connection matrix corresponding to possible interactions that have not yet been reported explicitly by neuroanatomists. The grossest possible perturbation of the data is simply to assume that all unreported connections exist (Young 1992). We analysed a matrix derived from coding all unreported conections as existing, and reflecting the resulting matrix about its leading diagonal in the manner described in §2*b*. Figure 6 presents the configuration that resulted from tied NMDS analysis of this matrix (see also Young 1992).

The structure in figure 6 is in many respects similar to that in figure 5. In both structures, the parietal and IT areas are strongly segregated, with V1 and some of the prestriate areas between them at the one side, and STP and area 46 between them at the other. In quantitative terms, the relation between the two structures is characterized by 76% variance-explained in the one structure by the other (Young 1992). The structure in figure 6 implies the same gross organizing principles as the structure in figure 5. Hence even when the grossest possible perturbation is applied to the data set, the conclusions are not disturbed. The solutions are similar because a large enough number of connections have been confirmed absent (see Felleman & Van Essen 1991). The principal differences between figures 5 and 6 lie in the shifts of position of less well-studied areas, such as VOT and V4t. These areas have their positions shifted toward the centre of the structure by their acquisition of a large number of hypothetical new connections, corresponding to possible connections that have not yet been explicitly ruled out. It would be surprising if these structures were in reality to possess such rich connectivity. Indeed, the positioning of these structures in the centre of the structure implies that the distribution of proximities in the connection matrix is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

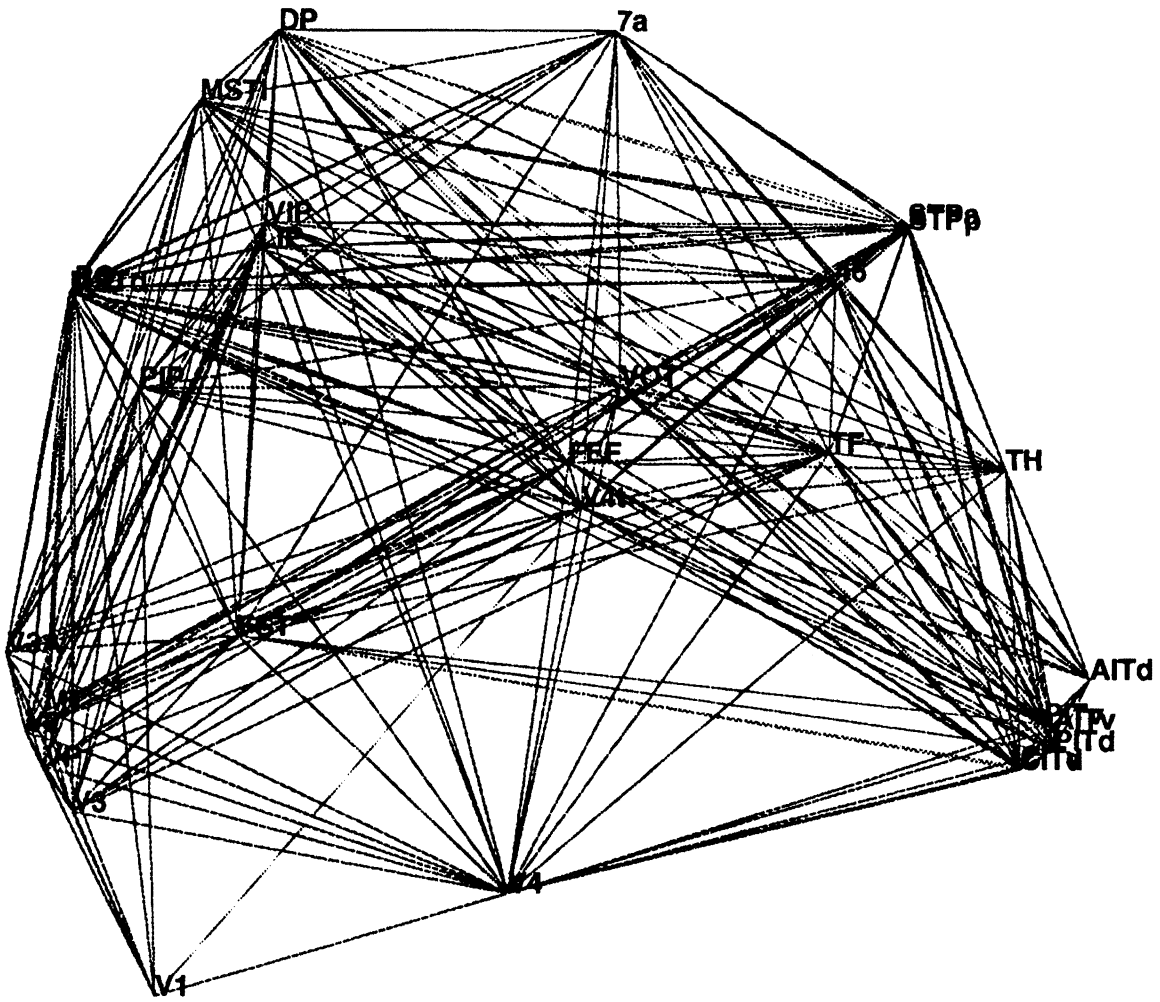

Figure 6. The structure derived from tied NMDS analysis of a visual system matrix in which all possible connections that have not been explicitly reported either present or absent have been assumed to exist. The structure still shows the gross features of figure 5, namely strong segregation between parietal and inferotemporal areas, a dimension passing between V1, at one extreme, and STP, at the other, and the positioning of area 46 and STP at the confluence of the two streams. The quantitative relation between this structure and that in figure 5 is characterized by 76% variance-explained. The principal differences between the two structures is the movement of some less well-studied areas, such as VOT and V4t, to the centre of the space because they have acquired a large number of hypothetical new connections. It is most unlikely that these areas really possess this rich connectivity, but the structure represents an example of the 'worst case', in which the data have been perturbed to the maximal possible degree.

very unlikely to reflect the underlying quantitative distribution of connection strength, for the reasons discussed in §2*a*,*c*: there are too many high and moderate proximities in this interpolated matrix. It follows that the assumption that all unreported connections exist considerably overestimates the real situation. In any case, the similarity between figures 5 and 6 suggests that the organizational conclusions are workably robust against changes in status of the possible connections that have not so far been reported.

The second 'control' analysis concerns whether different organizational principles would emerge from an untied NMDS analysis. As discussed in §3*b*, connection strengths are distributed continuously, with the exception of absent connections for which there is no meaningful underlying continuum. We argued that the tied approach was the appropriate one in the case of connection data. The tied approach treats proximities of the same value as members of discrete distributions. We considered this appropriate because the zero entries are truly discrete, and there is no basis (other than further neuroanatomy) for estimating for non-zero connection strengths where particular connections lie in the continuous distribution of strength. However, by removing the constraint that all tied proximities be reflected in distances that are as similar as possible, distances in an untied solution will vary continuously and this property of the solution will cause it to underreflect the segregation between particular points implied by zero proximities between them. Annular test data, for example, are represented as more disk-like than their real configuration in untied solutions. Hence, an untied solution can provide a lower bound estimate of the state of the cortical areas, when they are segregated to the least possible degree. This is an interesting estimate: if, for example, the parietal and IT areas are not segregated in such a solution, then caution would be indicated in concluding that there are two visual streams: the finding would be dependent on the discreteness of the distribution of distances in the tied solution. Accordingly, we undertook an untied NMDS analysis of the visual system data.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

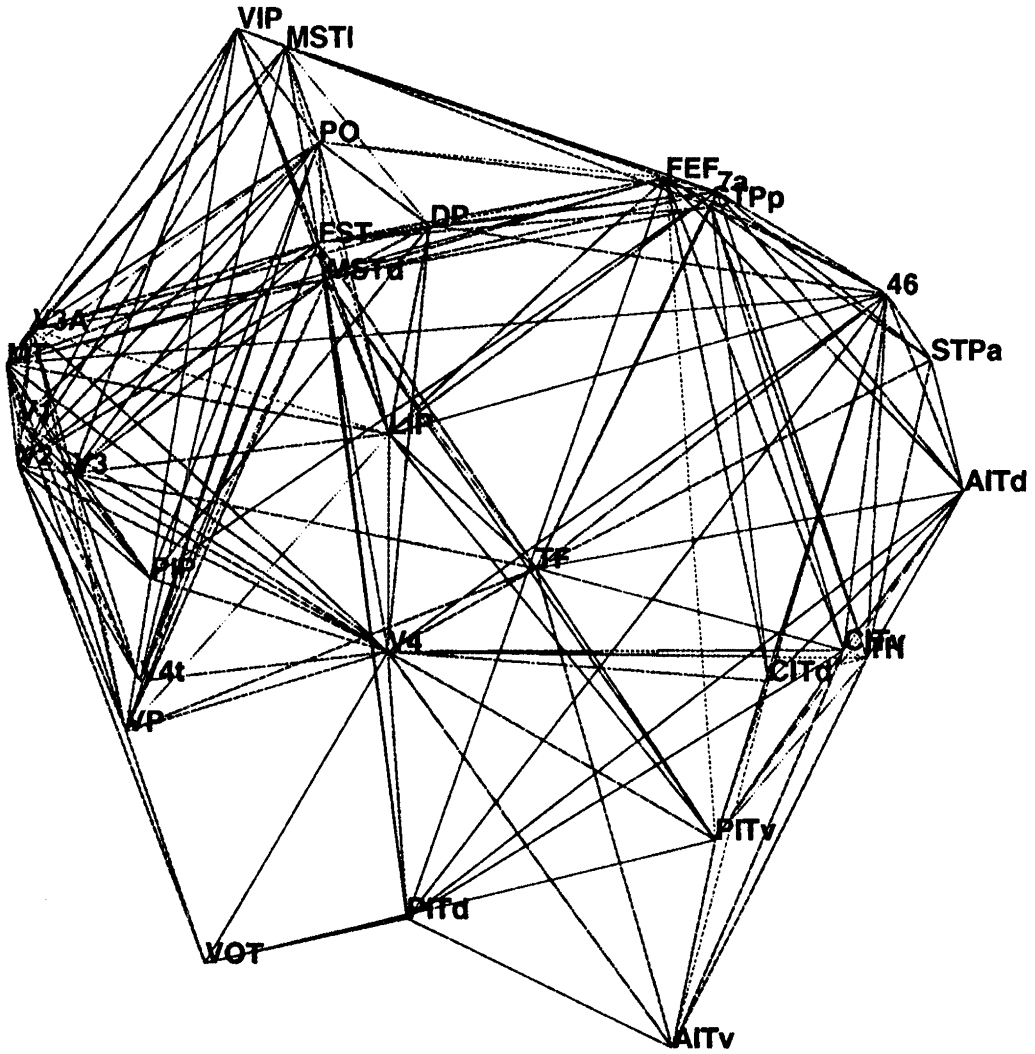

Figure 7. The structure derived from untied NMDS analysis of the visual system matrix. Untied NMDS analysis of structures with a very similar quantitative distribution of interpoint proximities to that of connection data (see §2*c*) understate the true separation of the points, by allowing points to stray into the centre of the structure. The structure therefore shows the optimal configuration when NMDS is allowed to under-represent the distances between areas. Nonetheless, note that all the dorsal stream areas are placed toward to top of the structure, that all the ventral stream areas are placed toward the bottom, that the hierarchical dimension from V1 to STPa and area 46 is present, and that STPa and areas 46 are again placed at the confluence of the two streams.

Figure 7 shows the untied solution for the visual system matrix. All the parietal areas are unequivocally placed in the top half of the structure, while all the IT areas are unequivocally placed in the lower half. The areas of the two streams are joined at the one side by V1 and the prestriate areas, and at the other by STP and area 46. Hence, even when violence is done to the properties of the data in the way described above, the conclusions about the organization of the system stand.

The third 'control' analysis is related to the question of whether the NMDS solution for the visual system data systematically reflects the data structure of the matrix. This question arises because where the STRESS of an NMDS solution is very high (or very low), an annular structure can emerge in the solution for spurious reasons. Randomly ordered data, for example, give rise to very high STRESS, and to annular configurations in which the ordering of the points within the annulus is random. The traditional means of deciding between the 'systematic-annularity' and 'spurious-annularity' cases is to compare the STRESS of a solution with analyses of comparable random data, thereby determining whether the solution is one in which very low fit could have produced artefact (Stenson & Knoll 1969). Accordingly, we examined the distribution of SSTRESS statistics from scaling 200 randomly re-ordered matrices. The probability of the SSTRESS of the solution in figure 5 falling within the distribution of random data with very low fit was less than 10^{-30} (Young *et al.* 1994). We performed an analogous analysis for the untied solution. The probability of the SSTRESS of the solution in figure 7 falling within the random data distribution was less than 10^{-11}. Hence these analyses provide no grounds for belief that the NMDS solutions for the visual system data should not be trusted to be a systematic reflection of data structure.

5. INDEPENDENT ANALYSES OF THE CONNECTIONAL ORGANIZATION OF THE PRIMATE VISUAL SYSTEM

Some aspects of the organization of the visual system have been, and remain, disputed. These include the question of whether there are discriminable streams within the system (Ungerleider & Mishkin 1982; Goodale & Milner 1993; Merigan & Maunsell 1993; cf. Martin 1992; Goodhill *et al.* 1994), whether there is any reconvergence of processed visual information (Perrett & Oram 1995; cf. Engel *et al.* 1991), and whether the system is a serially ordered hierarchy (Maunsell & Van Essen 1986; Felleman & Van Essen 1991; Young 1992; cf. Goodhill *et al.* 1994; Simmen *et al.* 1994). The results of NMDS analysis of the connectivity of the visual system lend unequivocal support to one side in each of these disputes. Where there is dispute it is important to bring multiple methods, involving different types of analysis and preferably other types of data, to bear on the various issues. In this way, opportunities for independent corroboration or discorroboration of results are provided. In the case of discorroboration between independent analyses, dispute will doubtlessly continue with the apologists for opposing views each emphasizing the aspect of the inconsistent results that supports their position. Should different analyses corroborate one another, however, then it is usual scientific practice that the independently verified finding either be accepted by both parties, or that an explanation is offered of how several methods can come independently to the same conclusion if it is not the correct one. We therefore now turn to independent analyses of the organization of the visual system, and the relation that their results bear to the results of the application of NMDS to the visual system connection matrix.

(*a*) *Simple statistical properties of the connection pattern*

The simplest means of testing some of these organizing principles is to employ a χ^2 test to determine whether the incidence of connections is that expected under the various hypotheses. One important dispute concerns whether the visual system is segregated into streams. Both analysis of the behavioural effects of cortical lesions (Ungerleider & Mishkin 1982) and NMDS (Young 1992) (and, to an extent, the physiological evidence; Merigan & Maunsell 1993) suggest strongly that the system is split into streams. However, it is sometimes maintained that the system is not internally segregated into streams (Martin 1992; Goodhill *et al.* 1994). These hypotheses are simple to test decisively. If the system is dichotomized, then the elements of the dorsal stream should be significantly more connected with their associates than with the elements of the ventral stream, and vice versa. Similarly, there should be significantly more connections that have been confirmed absent between dorsal and ventral areas than within each of these groupings. If these comparisons fail to reach significance, then the Null hypothesis that the areas are not segregated could not be rejected. Exactly comparable analyses can bear on the issue of whether the visual areas are serially ordered.

We divided all the areas into four sets, which corresponded to 'early', 'late', 'dorsal' and 'ventral' groupings. The 'early' set contained V1, V2, V3, VP, V3A, PIP and V4t; the 'late' set contained areas FEF, 46, STPa, STPp, TF and TH; the 'dorsal' group contained MT, MSTD, MSTL, FST, PO, LIP, VIP, DP and 7a; the 'ventral' group contained V4, VOT, PITD, PITV, CITD, CITV, AITD and AITV. Turning first to the analysis of demonstrated connections, the dorsal and ventral areas are much more connected internally than the Null hypothesis predicts ($\chi^2 = 17.2$, $p < 0.00004$). Turning to the analysis of connections that have been demonstrated absent, the dorsal and ventral areas exchange many fewer connections than the Null hypothesis predicts ($\chi^2 = 18.6$, $p < 0.00002$). The hypothesis that the visual areas are distributed without segregation is rejected. These results therefore contradict the suggestions of Martin (1992) and Goodhill *et al.* (1994) about the organization of the visual system, and support the findings from analysis of empirical data (Ungerleider & Mishkin 1982; Young 1992).

We now address the issue of whether the visual areas are serially ordered. If this hypothesis is true then areas at the bottom of the series should be significantly more connected with their associates than with areas at the top of the series. Similarly, there should be significantly more confirmed absent 'connections' between early and late visual areas than within each of these groupings. If these comparisons were to fail to reach significance then the Null hypothesis that high and low areas are not segregated could not be rejected. Early and late areas are significantly more connected internally within each group than would be expected on the Null hypothesis ($\chi^2 = 9.8$, $p < 0.002$). Further, early and late areas are significantly less connected between each other than would be expected on the Null hypothesis ($\chi^2 = 10.9$, $p < 0.001$). The Null hypothesis is thus rejected in this analysis also: early and late visual areas are segregated.

The analyses of the confirmed absent entries in the connection matrix also illustrate the fact that important information about the organization of the system is carried by absent connections, and therefore that it would be helpful if these absences were more often reported explicitly.

(*b*) *Procrustes rotation and approximate randomization*

A useful means of examining the relations between the results of different analyses that bear on the same objects is to employ Procrustes rotation (Schonemann & Carroll 1970; Gower 1971; Young 1990). In many cases, the dimensions on which the coordinates of the points are defined are not themselves meaningful, and it is the inter-relationships between the points that represent information about the objects. Here Procrustes rotation can be used to find the reflection, translation, rotation and scaling of one structure that fits another structure as well as possible. The optimal transform is found that minimizes a simple

sum of squared distances measure between the corresponding points of the two compared structures. The goodness-of-fit of the two structures can be expressed by the cumulative residual sum of squared distances between points (r.s.s.), or by a variance-explained statistic (1-r.s.s, or r^2), as we have preferred (Young 1992; Scannell *et al.* 1995).

The statistical rarity of each comparison effected by Procrustes rotation can be assessed by an approximate randomization test (Edgington 1980; Young 1990). This process repeats the rotation with one of the two compared structures shuffled randomly on each of a large number of iterations. The number of iterations depends on the amount of computer power available. In earlier reports we have used only 600 iterations, but we are now able to perform 10^6 shuffled rotations. The number of times that the variance-explained statistic is exceeded during these randomized iterations is divided by the number of iterations to yield a probability that a correspondence as good as the empirical, unshuffled, comparison could have come about by chance (Edgington 1980).

(*c*) *Seriation*

A non-NMDS method, normally used to examine the chronological order of Archaeological grave site data (Wilkinson 1971; Laporte & Taillefer 1987), has been suggested by Simmen *et al.* (1994) as a means of assessing serial ordering in connection data. This analysis is related to computational methods that undertake the travelling salesman problem by the common task of finding the unidimensional ordering of the elements of a matrix among the $n!$ possible orderings that minimizes some measure of distance between elements. In the case of Archaeological data, seriation algorithms have often been applied to incidence matrices, whose columns correspond to different kinds of objects (e.g. pottery), whose rows correspond to the graves in which the objects are found, and whose entries indicate whether a particular kind of object was found in a particular grave (Laporte & Taillefer 1987). The algorithm finds the permutation of the rows which minimizes the cumulative mismatch between all rows. The row order then specifies the serial order of the graves. Two types of information can emerge from the analysis. The optimal serial ordering of the graves is naturally the principal result, but it is also possible to derive a measure of departure from perfect serial ordering by comparing the cumulative number of mismatches between all rows, the circuit length, with theoretical minimum and average random circuit lengths.

In the case of connection data, the task of the seriation algorithm is again to find the permutation of the rows and columns with shortest circuit length, the only major difference being that the connection matrices are n by n. Just as with grave site data, the row ordering is the principal goal of the analysis since this specifies the optimal serial ordering of the cortical areas. Similarly, the optimal circuit length can be compared with the theoretical minimum ($2n$) and average random circuits to assess departure from perfect serial ordering. Unlike Archaeological data, however, it need not be assumed that there are specific start and finish points in the sequence, and the ordering can be circular (Hubert 1974; Simmen *et al.* 1994). We implemented a seriation algorithm using simulated annealing, and applied it to the connectivity matrix for the visual system (Young *et al.* 1994).

The shortest circuit length found by the algorithm in 10^6 trials was 196. Unfortunately, the program found not one but 60 different orderings at this optimal length. This particular application of the method hence suffers from the problem of multiple minima. The orderings differ in rather minor ways, however, mainly in a number of pairwise swaps between nearest neighbours in the orders. Several features were common to all the optimal length orderings. Parietal and IT areas were maximally segregated, being joined at the one side by V1 and the prestriate areas, and at the other side by area 46 and the parcellations of STP. Three randomly selected orderings are depicted in figure 8, in which both the orderings and the city-block distances between adjacent rows are represented, the latter being shown as the proportion of the circumference of the circle defined by the quotient of the distance between rows and the total circuit length.

We used Procrustes rotation to provide a quantitative measure of the relation between the results of the seriation algorithm and that of NMDS. Comparison by Procrustes of the arrangement of areas in each of the 60 ordering with the NMDS solution (figure 5) yielded 60 correlation coefficients distributed about a mean of 0.9 with standard deviation 0.006 (every $p < 0.000001$). Hence, this seriation method presents a conclusion that is not only qualitatively similar to that from NMDS (because the same organizing principles emerge from it) but is also strongly associated with the NMDS conclusion at a quantitative, statistically significant level. The strong relation between the results of these two independent analyses corroborates the results of each method. Because the high correlation between these independently derived results would almost never be expected by chance, disputation of the organizing principles that emerge from both analyses should include an account of how they can both arrive at the same conclusion if it is not the correct one. In particular, methodological artefact in one or the other of the methods does not explain this mutual corroboration (Young *et al.* 1994; cf. Simmen *et al.* 1994), as artefact of precisely the same form would be required to afflict both radically different methods to bring about the correspondence in their results.

The second type of information that can emerge from seriation analysis is an indication of whether the data are, or are not, serially ordered. This may be derived by comparing the data's optimal circuit length with the theoretical minimum circuit length, and with the circuit lengths for randomly permuted data matrices (Wilkinson 1971; Laporte & Taillefer 1987; Simmen *et al.* 1994; Young *et al.* 1994). We turn first to examining the relation between circuit length and theoretical minimum tour length for these data. A useful means of determining this relation for a data set is to calculate the quotient of the circuit length divided

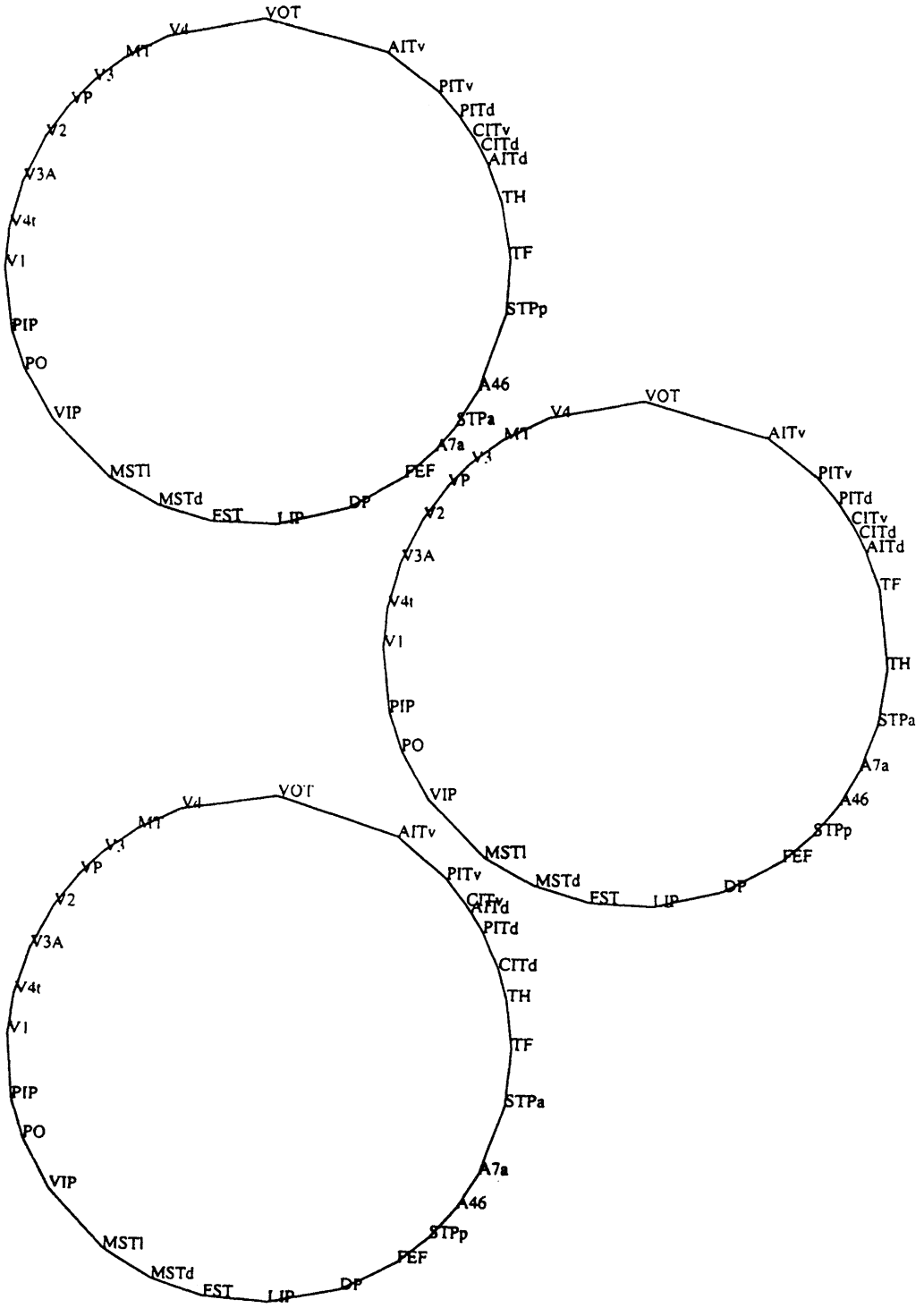

Figure 8. Three of the optimal orderings from seration analysis of the visual system matrix. V1 is placed at left in all diagrams. The distances between labelled vertices are proportional to the City Block distance between the row vectors of each corresponding area. All the optimal orderings possessed the following features: parietal and inferotemporal areas were maximally segregated, and these areas were joined at the one side by striate and prestriate visual areas, and at the other by STP and area 46. The differences between the optimal orderings were minor, and consisted principally in pairwise swaps between nearest neighbours in the orderings. Areas TF and TH, for example, are swapped between the top and middle orderings.

by the theoretical minimum circuit length. The smaller is this quotient, the stronger is the serial ordering. The optimal circuit length for the visual system data was 196, and their theoretical minimum circuit length was 60. The quotient for these data was therefore 196/60 = 3.27. It is interesting to compare this quotient with comparable Archaeological data which are accepted as exhibiting serial ordering. For example, Archaeological data reported by Dempsey & Baumhoff (1963) (n = 26) give a quotient of 4.9, those from Kuzara *et al.* (1966) (n = 11) gives 3.86, Hole & Shaw (1967) (n = 57) 3.3, and Stefan (1971) (n = 49) 3.21. The relation between optimal circuit length and theoretical minimum for the visual system data therefore characterizes

data that are at least as strongly serially ordered as many empirical data sets in Archaeology that are accepted as possessing serial order.

Another comparison of interest is that between the optimal circuit length and the distribution of circuit lengths for permuted data (Simmen *et al.* 1994; Young *et al.* 1994). The values of the mean circuit length for 300 permuted data matrices and the standard deviation of the distribution were 420 and 15, respectively. Conventional statistical inference suggests that actual circuit lengths shorter than 2 standard deviations of the mean permuted tour length should cause the Null hypothesis to be rejected at the 0.05 level, compelling the conclusion that serial ordering is present. According to these values, any tour of length less than 390 should then compel the conclusion that serial order is present by this conventional criterion. The actual value for the visual system data's optimal circuit length, 196, is 14.9 standard deviations short of the permuted data's mean. This corresponds to a significance level, assuming a Gaussian distribution of permuted data circuit lengths, of almost 10^{-50} (Young *et al.* 1994).

It is apparent that the visual system connectivity matrix is certainly serially ordered. However, the original purpose in suggesting this analysis was to discriminate between the 'segregated into streams' and 'no segregation' hypotheses about the organization of the visual system (Simmen *et al.* 1994). The rationale for this application was that the data should exhibit strong serial ordering under the former hypothesis, but not the latter (Simmen *et al.* 1994). Unfortunately, circuit length is not a sufficiently sensitive measure to bear on this issue. Even unsegregated data, in which points are randomly but uniformly distributed within a plane can possess short circuit lengths, simply because efficient unidimensional tours can be found through the points (see, for example, Mitchison & Durbin 1986). This is the routine application of travelling salesman algorithms. The original proposal for the use of this method assumes that a significantly shorter circuit will be found around the circumference of a ring than through the middle of the space that the ring encloses, but it is easy to see that this need not be the case. Hence even though the data are serially ordered, this fact does not inform the issue of whether there are internal streams within the visual system.

(*d*) *Hierarchical analysis*

A method for defining a hierarchical arrangement of a set of cortical areas based on the laminar origins and terminations of their connections was originated by Rockland & Pandya (1979). This method has since been developed by others (Maunsell & Van Essen 1983; Felleman & Van Essen 1991) by describing further criteria by which cortico-cortical connections can be classified as 'ascending', 'descending', and 'lateral'. Projections classified as ascending originate in the supragranular or supragranular and infragranular layers and terminate predominantly in layer 4 of their target. Lateral projections originate in a bilaminar pattern from both the supragranular and infragranular layers and terminate throughout the thickness of the cortex, in a columnar fashion. Descending projections originate bilaminarly or in the infragranular layers and terminate in the superficial supragranular and/or deep infragranular layers. These rules for classifying connections, together with data on the laminar origin and termination of connections, can be used to define a table of hierarchical constraints. A hierarchical ladder can then be derived by arranging the cortical areas among an arbitrary number of levels in such a way that the hierarchical ordering of the areas matches the table of hierarchical constraints as well as possible. Relations between hierarchical ladders derived in this way with the results of NMDS are particularly interesting, since this approach involves not only an independent method but also a completely different type of data.

We have constructed a numerical model from the most recently published hierarchical ladder (Felleman & Van Essen 1991). The model was constructed by associating an integer value with each visual cortical area according to its height above V1 in the Felleman & Van Essen (1991) scheme. We then used Procrustes rotation to compare this hierarchical model with the configuration that emerged from NMDS analysis of the visual system connectivity matrix (Young 1992). The hierarchical model, which posseses only one dimension, accounted for 30% of the variability of the two-dimensional NMDS structure, and this relation was, according to an approximate randomization test with 10^6 iterations, statistically significant at $p < 0.000001$. These completely independent methods, based on completely different types of data, therefore concur about the ordering of cortical stations in the visual system to a degree that their agreement would almost never be expected by chance. Using the now familiar logic, disputation of the conclusions of either NMDS or hierarchical analysis should explain how such remarkably close corroboration could come about by chance or artefact, if not by two independent analyses of different types of data faithfully reflecting the same underlying organization.

(*e*) *Analysis of the behavioural effects of cortical lesions*

One of the contended principles of organization of the visual system has been the proposal, made most clearly by Ungerleider & Mishkin (1982), that the system is divided into distinct dorsal and ventral pathways. The basis for this proposal was the pattern of behavioural deficits in monkeys which had received intelligently placed cortical lesions. Damage to the temporal cortex produces impairment in visual recognition, whereas damage to the parietal cortex produces a constellation of visuo-spatial impairments (Ungerleider & Mishkin 1982). A great deal of anatomical and physiological work has been undertaken in the intervening decade, a proportion of which has been directed specifically at this idea, but it remains a working hypothesis for very many visual cortical researchers. One recent development has been the refinement of the cortical parcellation to which the

earlier research had access. The task of rendering the organizational principle described by Ungerleider & Mishkin (1982) into a numerical model therefore requires assigning each of the cortical areas of the more modern parcellation system that informed the NMDS analyses, into a group of areas that would be recognizable to the proponents of this idea. Accordingly, we assigned any area of the parietal cortex, any area of the caudal superior temporal sulcus, and any area whose function is largely oculomotor (e.g. FEF) to the 'dorsal' group. Analogously, we assigned all ventral areas, including areas of the inferotemporal and parahippocampal cortices and V4 to the 'ventral' group. To capture the common origin of the streams in occipital cortex, we assigned appropriate areas to the 'shared' group (e.g. V1, V2). We assigned the value 1 to all the ventral stream areas, 2 to the shared areas, and 3 to the dorsal stream areas. Hence, a simple unidimensional model with only three points captured the organizational principles taken from the behavioural effects of cortical lesions.

We then used Procrustes rotation to compare this model with the NMDS structure derived from analysis of visual cortical connectivity data (without STPa and area 46, since the Ungerleider & Mishkin model did not anticipate reconvergence of their streams in these structures). The model, despite its simplicity, accounted for 56% of the NMDS structure. Approximate randomization showed that a correspondence as close as this would not appear on more than one occasion in one million by chance ($p < 0.000001$). Again, therefore, independent analyses of different kinds of data bearing on the organization of the visual system exhibit mutual corroboration. An account of the visual system that holds that it is not divided into streams should therefore account for the fact that multiple independent methods corroborate one another in concluding that the system is dichotomized.

The results of all the different methods of analysis that bear on the organization of the visual system are summarized in §7.

6. ARE ANY ASPECTS OF CONNECTION DATA STRUCTURE OBSCURED BY SPARSITY?

The shape of the manifold in which the cortical areas are distributed is indicated by the distribution of quantitative connection strengths (§2*c*). This manifold takes the form of a ring or horseshoe in which the points representing the cortical areas are placed optimally by NMDS. The curvature of the structure is a necessary consequence of the many long distances implied by the many zero proximities in quantitative connection data and their qualitative representations in connection matrices. We have seen that the curvature that attends sparsity is a reflection of a *bona fide* aspect of data structure, but it is one that sometimes does not give much information beyond that calculable from the simple proportion of zero entries in a connection matrix. The manifolds' gross shapes, moreover, do not emphasize the influence of the non-zero entries, and their curvature can occupy a dimension of the solution space. Because humans find it difficult to apprehend spatial relations in more than three dimensions, and not all aspects of connection data structure may be fully represented in the low number of dimensions that can be apprehended, we have explored methods of diminishing the effects of sparsity on the output configurations. Our purpose in pursuing these methods is to ensure that as many aspects of data structure as possible, particularly those aspects of data structure that are carried by non-zero entries, can be visualized in NMDS solutions. Our general methodology in this section is to interpolate the zero entries on the basis of the non-zero ones, to reduce the influence of the former.

(*a*) *Data conditioning methods*

Kendall (1971) described a data conditioning procedure by which the horseshoe produced by NMDS analysis of a sparse, low-level matrix (containing Archaeological data) may be unbent. Many methods for data conditioning and data transformation have since been described (Kendall 1975; Gower & Legendre 1986; Lefkovitch 1991). We have adopted, modified or been inspired by a number of these methods which seemed to us particularly suitable. To explain the nature of the methods we have used, we consider a sparse matrix **M** with elements m_{ij}, whose level of measurement is ordinal and whose number of distinct levels of proximity is low, which is transformed into a matrix **S** with elements s_{ij}, whose level of measurement is also ordinal but whose number of distinct levels of proximity is higher, and whose proportion of zero elements is much reduced. We then describe the transformation that each method performs. The data conditioning methods fall into two broad classes, the first being concerned with the similarity or dissimilarity of the pattern of connections between any two cortical areas, and the second being concerned with the number of steps that need to be taken between unconnected cortical areas, to reach the one area from the other.

(*i*) *Similarity or dissimilarity of connection pattern*

The first data conditioning transform is simply to sum the absolute differences between any two rows, and to treat this sum as the new measure of proximity between the rows. In this case, a new matrix **S** with elements s_{ij} is derived

$$s_{ij} = \sum_k (m_{ik} - m_{jk}).$$

We termed this method 'dsm', and treated its output **S** as ordinal dissimilarity matrices in their subsequent NMDS analysis.

The further methods in this section were inspired particularly by the vector dissimilarity technique described by Lefkovitch (1991). The second method involves a weighted dissimilarity transform, where the weighting is related to the total number of non-zero entries in each comparison:

$$s_{ij} = [\sum_k |(m_{ik} - m_{jk})|] / [\sum_k |(m_{ik} + m_{jk})|].$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

This method was termed 'wdsm1', and devised to accord greater dissimilarity to vector comparisons that contain more zero entries. As with most transforms (see Kendall 1971), this procedure can be applied repeatedly, to derive further higher-order matrices. We derived a third set of matrices by re-applying this transform to the wdsm1 matrices, the resulting matrices being termed 'wdsm2'.

The fourth method was designed to capture the similarity between two vectors' non-zero entries. It did not regard common zeros as indicative of similarity. The transform was

$$\text{if } m_{ik} > 0, \text{ and } m_{ij} > 0, \; s_{ij} = \sum_k (m_{ik} + m_{jk}).$$

We designated this method 'nzsim1', and by re-applying the method to the output nzsim1 matrices, derived a fifth set of matrices 'nzsim2'.

The sixth set of conditioned matrices were derived by a combined similarity–dissimilarity transform. The similarity between two vectors was derived by nzsim1, and the dissimilarity by the dsm procedure. The values s_{ij} in the output matrices were then derived by subtracting the latter from the former. This set of matrices were termed 'simdsm' matrices.

(ii) Path length

The data conditioning methods in this section were adapted from Kendall's method of evaluating pathlength (Kendall 1975). Consider the case that

$$m_{ij} = m_{kl} = 0.$$

In this case, d_{ij} and d_{kl}both represent distances above a threshold where the actual distance is not known. In order to compare $m_{ij}$with m_{kl} we consider how one might leapfrog from the ith to the jth object through intermediate objects by the shortest possible route, and compare the length of that path with that joining the kth and the lth object. A concrete example is that of thresholded distances between cities (see §6*c*). Let us assume that Aberdeen to London and Birmingham to London possess zero similarity because both distances are above an arbitrary threshold. We try to resolve the ambiguity between these large distances by using the fact that there is a shorter pathlength between Birmingham and London than between Aberdeen and London. In the former path, both London and Birmingham could have non-zero similarity with Oxford (i.e. pathlength 2 between them), whereas many more steps would have to be taken to link Aberdeen and London, indicating that they are farther apart.

We define the qth order pathlength matrix $\mathbf{P}^{(q)}$ with elements $p_{ij}^{(q)}$ where

$$p_{ij}^{(1)} = \max\,(m_{ix} m_{jx}),\; x = 1, 2, \ldots, n$$

and

$$p_{ij}^{(2)} = \max\,(m_{ix} m_{yx} m_{jy}),\; x = 1, 2, \ldots, n$$

or, more generally, the qth order pathlength matrix is

$$p_{ij}^{(q)} = \max\,(m_{ix1}\, m_{x1x2}\, m_{x(q-1)xq} m_{xqj}),$$
$$x_1, x_2, \ldots, x_q = 1, 2, \ldots, n.$$

We examine only the lowest order pathlength correction for each pair of objects: that is, if a qth order path exists between objects i and j we do not further consider paths that involve more than q intermediate objects.

We assume that if

$$p^{(q)}{}_{ij} > p^{(q)}{}_{kl},$$

then

$$m_{ij} > m_{kl}$$

and that if

$$p^{(q)}{}_{ij} > 0,\; p^{(q)}{}_{kl} = 0 \text{ and } p^{(q+1)}{}_{kl} > 0,$$

then

$$m_{ij} > m_{kl}.$$

We implemented this procedure by defining a matrix of proximities $\mathbf{S}$ with elements s_{ij} such that

$$s_{ij} = e_0 + e_1 p_{ij}^{(1)} + e_2 p_{ij}^{(2)} + \ldots + e_q p_{ij}^{(q)}$$

where e_q is a scaling factor to ensure that latter condition is satisfied and e_0 ensures that the extra ranks inserted into the corrected matrix do not overwrite the ranks that exist in the original data;

that is,

$$e_q p_{ij}^{(q)} > e_{q+1} p_{ij}^{(q+1)}\; \forall q, i, j$$

and

$$e_0 > \Sigma e_q p_{ij}^{(q)}\; \forall q, i, j.$$

We termed the matrices that emerged from the application of this method 'pth1'. We defined a second pathlength transform, in which the criterion by which we select the best path involves and additive operator rather than the multiplicative one used in pth1. In this second general case the qth order pathlength matrix is

$$p_{ij}^{(q)} = \max\,(m_{ix1} + m_{x1x2} + m_{x(q-1)xq} + m_{xqj}),$$
$$x_1, x_2, \ldots, x_q = 1, 2, \ldots, n,$$

where exactly comparable scaling factors and assumptions as in pth1 apply. We termed the matrices from this method 'pth2'.

(*b*) *Recovering metric parameters from test data at the same level of measurement as connection data: a tournament*

We examined quantitatively the performance of the data conditioning and transformation techniques described above. The object of this examination was to determine optimal NMDS approaches to the many aspects of structure in neuroanatomical connection data, and to ensure that the data's sparsity does not obscure aspects of data structure that are not carried by the zero entries. Accordingly, we produced a large number of arbitrary test structures, and derived several types of proximity data from them. First we derived Euclidean distance measures between the vertices of these structures. Second we lowered the Euclidean distance measures to the same level of measurement as neuroanatomical connection data. For the 'connection-type' data, we established two different

Table 1. *This table summarizes the goodness-of-fit statistics (r.s.q.) from Procrustes rotation in the data conditioning tournament.*

(Each of the data transformation routines are enumerated across the top of the table, and ring and disk data structures are listed at the left. The 0, 1, 2, and 0, 1, 2, 3, 4, 5, 6 rating scale results are indicated separately, as are the tied and untied results for each type of structure. Each r.s.q. statistic is associated with its standard deviation (across the 50 structures of each type). Each r.s.q. value may be rendered into a correlation coefficient by taking its square root: for example, the untied wdsm1 analysis of ring data structures represented by a 0, 1, 2 rating scale gave a r.s.q. of 0.958, and therefore a correlation of 0.98.)

		nzsim1	nzsim2	wdsm1	wdsm2	simdsm1	simdsm2	untrans	pth1	pth2
Ring, 0-2										
tied	mean rsq	0.951	0.549	0.954	0.881	0.947	0.918	0.955	0.954	0.935
	stdev	0.019	0.142	0.018	0.048	0.019	0.028	0.016	0.015	0.131
untied	means rsq	0.949	0.556	0.958	0.882	0.947	0.916	0.946	0.952	0.952
	stdev	0.021	0.141	0.015	0.047	0.019	0.029	0.015	0.017	0.017
Ring, 0-6										
tied	mean rsq	0.944	0.764	0.945	0.869	0.931	0.871	0.943	0.948	0.949
	stdev	0.016	0.083	0.017	0.048	0.022	0.044	0.018	0.022	0.022
untied	mean rsq	0.947	0.765	0.953	0.873	0.931	0.872	0.931	0.949	0.949
	stdev	0.016	0.083	0.012	0.045	0.022	0.044	0.024	0.022	0.022
Disk, 0-2										
tied	mean rsq	0.913	0.446	0.928	0.699	0.861	0.735	0.869	0.911	0.8931
	stdev	0.054	0.166	0.027	0.155	0.115	0.14	0.106	0.04	0.1334
untied	mean rsq	0.914	0.442	0.933	0.695	0.864	0.737	0.899	0.912	0.9124
	stdev	0.047	0.162	0.022	0.159	0.114	0.14	0.062	0.039	0.0392
Disk, 0-6										
tied	mean rsq	0.911	0.626	0.918	0.699	0.843	0.679	0.839	0.904	0.904
	stdev	0.046	0.135	0.028	0.142	0.098	0.135	0.158	0.043	0.043
untied	mean rsq	0.901	0.621	0.914	0.691	0.833	0.673	0.871	0.903	0.896
	stdev	0.054	0.139	0.046	0.154	0.115	0.139	0.081	0.043	0.043

sets of quantization thresholds by which the Euclidean distance matrix for each test structure would be transformed. In the first, the thresholds were placed so that the degraded matrices contained values of 2, 1 or 0, in the same proportions as those for the primate visual system data. In the second, we derived matrices with values 6, 5, 4, 3, 2, 1 and 0.

We then transformed each of the test structures' degraded matrices by each of the data conditioning methods described in §6*a*. The performance of each of the conditioning methods was tested, as was the performance of NMDS on the untransformed matrices. We achieved this by using Procrustes rotation to compare the NMDS solutions for Euclidean distance data with the corresponding degraded and degraded-and-conditioned matrices. Thus, the metric parameters of each test structure (as represented optimally in two dimensions), which presumably reflect all aspects of data structure, were systematically compared with the optimal solutions for each data conditioning method, and with unconditioned data at the same level of measurement as connection data. For example, a single test structure's metric solution would be compared with 18 other structures, namely the tied untransformed structure, the tied dsm, tied wdsm1, tied wdsm2, tied nzsim1, tied nzsim2, tied simdsm, tied pth1, tied pth2, the untied untransformed structure, the untied dsm, untied wdsm1, untied wdsm2, untied nzsim1, untied nzsim2, untied simdsm, untied pth1, and untied pth2 structures. The goodness-of-fit of each of these structures with the metric structure was returned by Procrustes rotation and used to assess the success of each in recovering the data's metric paramenters.

The test structures that we examined were of three types. We investigated a number of regular and geographical data sets, such as a plane in which 30 points were positioned in a regular hexagonal lattice, and a data set composed of the road distances between 30 British cities. We created 50 different test configurations in which 30 points were arrayed at random, but with a rectangular probability distribution, within a circular planar region. We further created 50 different test configurations in which 30 points were arrayed at random, with a square probability distribution, within an annular region of a plane. Both the disk and ring structures were perturbed by noise according to the standard treatment by Young (1970)

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

so that the mean SSTRESS of the tied NMDS analyses of these structures' degraded three-point rating scale data did not differ statistically from that for the NMDS analysis of the visual system data.

Analyses of some of these test structures are irrelevant to the question of recovery of parameters from cortical connection data. It is impossible, for example, for cortical connection data to specify regular grids or uniformly distributed disk structures because of the constraints imposed by the quantitative distribution of connection strengths for cortical data (see §2*c*). We have included these other structures for two reasons. Firstly, we wished to test some of the generality of the methods. Secondly, we included disk configurations because some subcortical brain structures may exhibit distributions of connection strength that are more

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

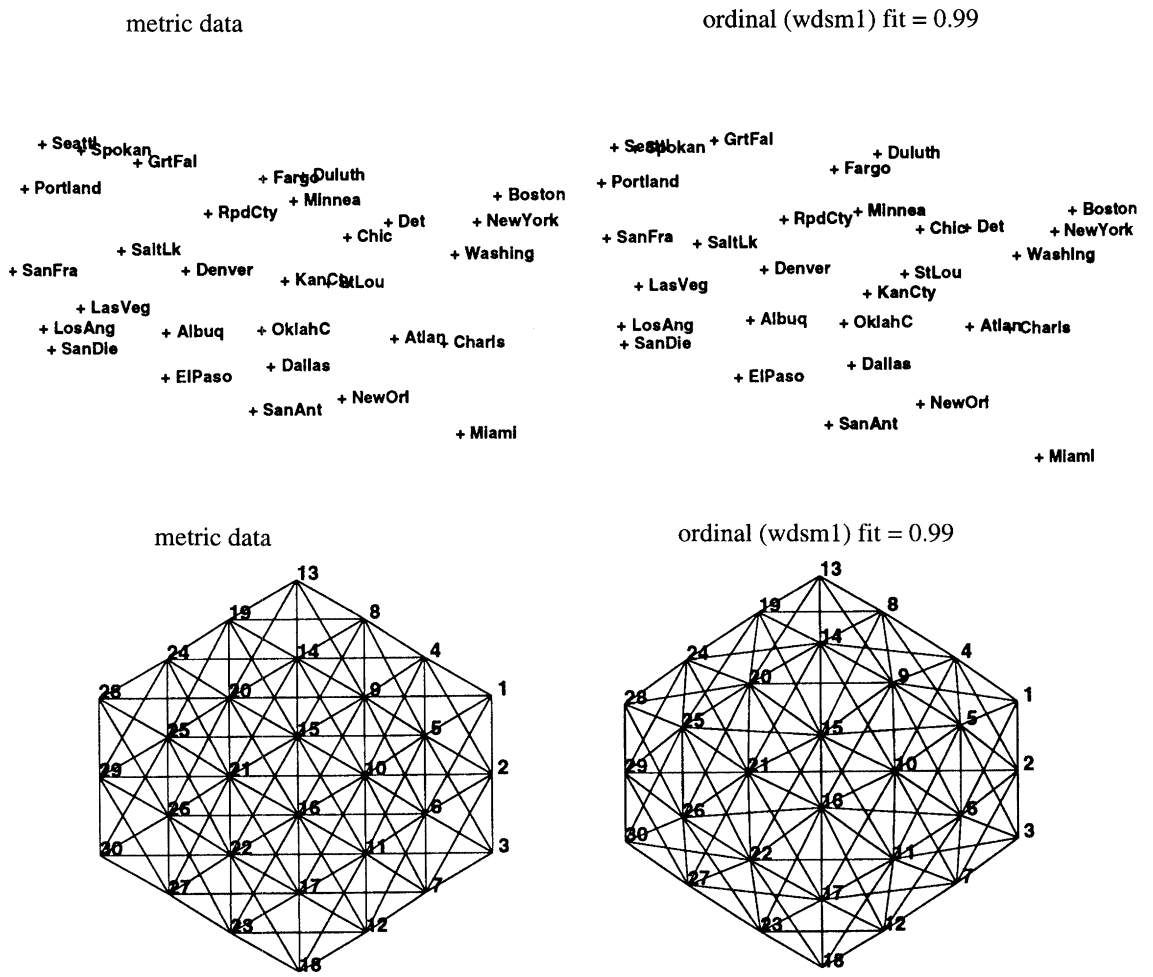

Figure 9. Examples of output from the successful methods of recovering metric parameters from matrices at the same level of measurement as connection data. On the left are the metric structures for distance data between 30 U.S. cities, and a regular hexagonal grid. On the right are the ordinal solutions for matrices derived using data conditioning method wdsm1. At the top right the positions of the cities are shown to have been recovered almost perfectly. The fit between the conditioned low level data and the metric data solutions for the cities was 0.99. At the bottom right is the solution for data treated by wdsm1 from the regular hexagonal grid. The fit between metric and conditioned data solutions was 0.99.

consistent with their being placed in the central region of a solution, so that these interpolation methods may be of value in the future when subcortical structures are analysed.

Table 1 summarizes the results of the data recovery tournament for the ring and disk test structures. All the methods did comparably well in recovering the structure of the data, except for two of those that involved repeated application of a transform (e.g. nzsim2 and wdsm2), and the unsophisticated dissimilarity transform (dsm). Apart from these poorly performing methods, the other treatments all recovered the data's parameters very well, their recovery typically being characterized by variance-explained statistics in the high 90% region. The successful methods (nzsim1, wdsm1, simdsm, untransformed, pth1 and pth2) did not differ significantly in performance, but method wdsm1 most often recovered the most variability, and method pth1 also performed notably well. Recovery of the parameters of disk data was slightly worse than that for the annular structures, but still averaged well above 90% for method wdsm1. Recovery of variability was slightly but consistently worse for the seven-point rating scale data than for the three-point rating scale. It is interesting to note that more proximity levels do not guarantee better recovery of the parameters of a data set (cf. Goodhill *et al.* 1994).

The ring structures, whose distributions of metric proximities were very similar to the distribution of connection strength for quantitative anatomical data (see §2*c*), were very well recovered by NMDS analysis. The average variance-explained statistic (across the 50 ring structures) for tied analysis of untransformed ring data, data which was at the same level of measurement and possessed the same proportion of each value in the rating scale as the visual system matrix, was 96%. This is comparable to the faithfulness of recovery of the Coombs and Kao configuration by Kruskal's NMDS algorithm, a recovery that was presented to demonstrate the basic power of the NMDS method (Kruskal 1964*a*). The reported analyses of neuroanatomical connection data are therefore likely to capture all but a tiny fraction of the structure inherent in the data (cf. Simmen *et al.* 1994), and are well within the envelope defined by remaining uncertainties about the experimental data. It is interesting that the untied

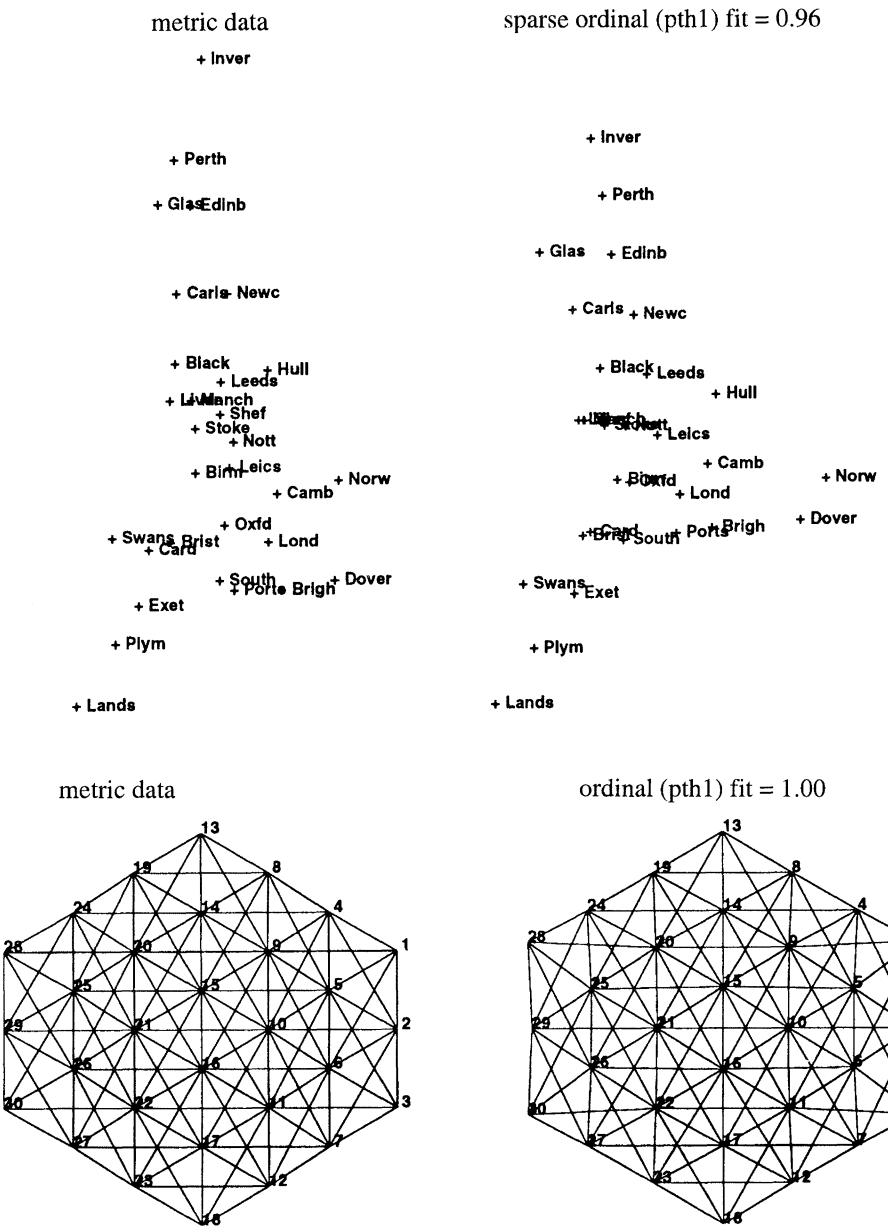

Figure 10. Examples of output from the pth1 method of recovering metric parameters from matrices at the same level of measurement as connection data. On the left are the metric structures for road mileage data between 30 British cities, and a regular hexagonal grid. On the right are the ordinal solutions for matrices derived using data conditioning method pth1. At the top right the positions of the cities are shown to have been recovered very well, despite the fact that the input data were set at the limit case for sparsity in which several cities had only one off-diagonal non-zero value. The fit between the conditioned low level data and the metric data solutions for the cities was 0.96. At the bottom right is the solution for data treated by pth1 from the regular hexagonal grid. The fit between metric and conditioned data solutions was perfect.

wdsm1 analyses were slightly more successful than the tied untransformed analyses, which are those that we have previously reported, but it is also interesting that the wdsm1 analyses were less than 0.3% better.

(*c*) *A demonstration of results from the successful data conditioning methods*

We now present some examples of the output from the most successful methods of recovering metric parameters from matrices at the same level of measurement as connection matrices (see figures 9 and 10). On the left of figure 9 are the metric structures for Euclidean distance data between 30 U.S. cities, and a regular hexagonal grid. On the right of the figure are the ordinal solutions for matrices derived by: (i) degrading the metric data to a three-point rating scale with the same proportions of 0s, 1s and 2s as the visual system matrix; and (ii) conditioning this low level data by method wdsm1. As can be seen at the top of the figure, the positions of the cities are recovered almost perfectly. The fit between the conditioned low level data and the metric data solutions for the cities was 0.99, showing that the wdsm1-NMDS method recovers: in practical terms, all aspects of structure in these data. On the bottom right of the figure is the solution for data treated by wdsm1 from the regular hexagonal

grid. Here the fit between metric and conditioned data solutions was 0.99.

We noticed that method pth1 was particularly successful in the analysis of matrices with sparsity at or near the limit case in which objects have only one off-diagonal non-zero entry. Figure 10 shows the performance of pth1-NMDS in a very sparse case. At the left of the figure is the metric structure for mileage data between 30 British cities. At the right of the figure is the reconstruction of the positions of the cities by pth1-NMDS from a three-point rating scale matrix of such sparsity that several cities (e.g. Inverness and Lands End) have only one 'connection' with another city. Even at the level of sparsity set to the limit case, the recovery protocol yielded a solution that accounted for 96% of the variability of the metric solution. The method recovered the regular hexagon perfectly.

These examples illustrate what was shown quantitatively in the preceding section: these data conditioning methods are capable of recovering all but a very small fraction of the metric parameters in a data set from data at a level of measurement and sparsity identical to neuroanatomical connection data.

(*d*) *Re-analysis of the visual system data with successful data conditioning methods*

We applied wdsm1 and pth1 to the visual system matrix. Both these data conditioning treatments are capable of reliably mitigating the synthetic sparsity that is introduced by thresholding metric distance data, as demonstrated in the preceding two sections. The visual system analyses are presented to ensure that as few aspects of data structure as possible are obscured by the data's genuine sparsity in the small number of dimensions required. Figure 11 shows the solutions derived by wdsm1 and pth1 for the visual system, with the solution at the top being that for wdsm1. Both solutions are remarkably similar, despite the very different algorithms by which the data were transformed. As can be seen clearly in both configurations, all the 'dorsal stream' areas are concentrated in the top part of each diagram, while all the 'ventral stream' areas are concentrated in the bottom half. The two streams originate in a number of occipital visual areas, including V1, which are placed at left, and appear to reconverge in STP and area 46, which are placed at the right.

To establish these features of the results with transformed data quantitatively, we used Procrustes rotation to compare the transformed-data structures with the model that explicitly captures the two streams conception of Ungerleider and Mishkin, with the untransformed-data structure (see figure 5), and with each other. The results were that the wdsm1 and pth1 structures both share 92% of their variability with the untransformed structure. The wdsm1 and pth1 structures share 93% of their variability with each other. They are both therefore very similar indeed, and very similar indeed to the untransformed structure. The two streams model explained 43% of the wdsm1 structure, and 42% of the pth1 structure. The probabilities of these correspondences coming about by chance, according to approximate randomization tests, are all less than one in a million. Hence, the same conclusions would be drawn concerning the gross organization of the system from these solutions as from the untransformed data.

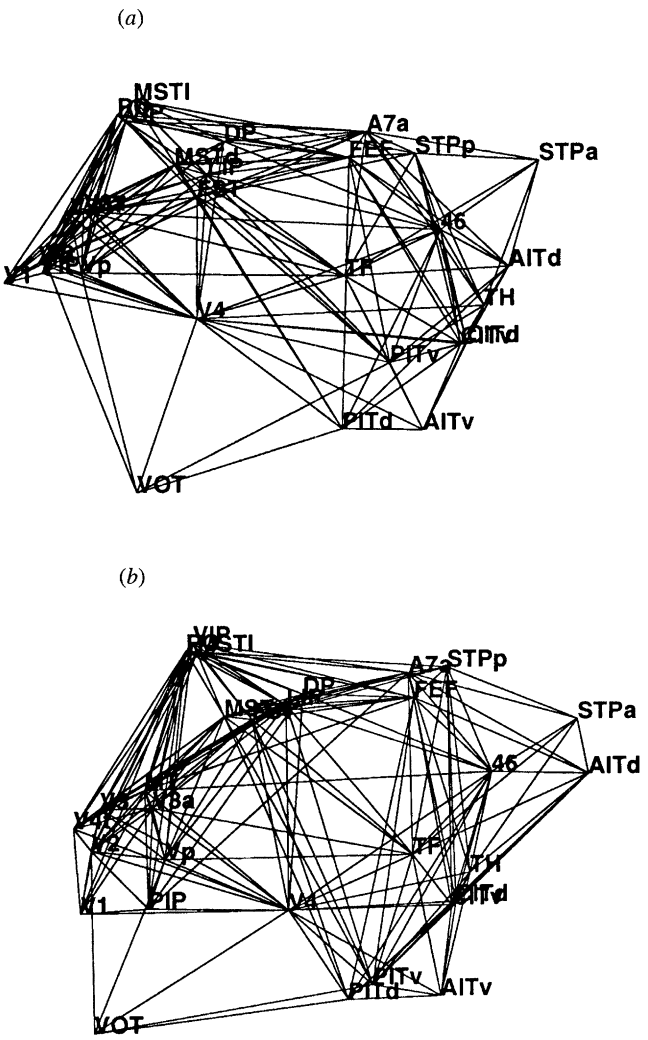

Figure 11. Structures derived by submitting the visual system matrix to (*a*) wdsm1 and (*b*) pth1 and then to analysis by untied NMDS. Both solutions are very similar, despite the very different nature of the algorithms by which the data were transformed. As can be seen clearly in both configurations, all the 'dorsal stream' areas are concentrated in the top part of each diagram, while all the 'ventral stream' areas are concentrated in the bottom half. The two streams originate in a number of occipital visual areas, including V1, which are placed at left, and appear to reconverge in STP and area 46, which are placed at the right. Precisely the same conclusions about the gross organization of the system would be drawn from these solutions as from the untransformed data analysis.

The two solutions, however, suggest that there may be further division of labour within the two visual streams. This is particularly apparent for areas in the dorsal stream, for which there seems to be further dichotomization which draws areas PO, MSTL and VIP away from their associates. Similarly, the possible distinction between V4 and TF and the other ventral stream areas, which is apparent in the untransformed tied solution (figure 5), is suggested even more strongly in these solutions. It is of note that, even when the effects of sparsity are mitigated or abolished, the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

configurations do not evidence an undifferentiated distribution of the visual areas, as some have suggested (Goodhill *et al.* 1994), but indicate a structure in which the dorsal and ventral stream areas are clearly discriminated.

7. HOW IS THE PRIMATE VISUAL SYSTEM ORGANIZED?

In this section we summarize, point by point, the results from NMDS analysis of neuroanatomical connection data, from analyses that bear on the reliability of NMDS analysis of this type of data, and from analyses which independently bear on the issue of the organization of the primate cortical visual system and which can therefore validate the NMDS analyses externally. We suggest that these facts constitute the principal explananda in this area, and we attempt to explain them in §8.

(*a*) *NMDS evidence*

1. Tied NMDS analysis of the visual system connection matrix, which we contend is the most appropriate, produces a solution of which four aggregate characteristics may be stated: the visual areas are dichotomized into two streams, both streams are serial hierarchies, the streams reconverge in high order areas, and neighbouring areas tend to exchange connections (§4; Young 1992).

2. In the global analysis of almost all cortical station in the Macaque, all the visual structures appear together at the left margin of the configuration (Young 1993), but the ventral and dorsal stream areas are unequivocally segregated, and a hierarchy of visual areas ascends from V1 to the rostral visual areas. Hence, two distinguishable hierarchically ordered streams are present in configurations that place all elements of the visual system on the same side of the NMDS output space.

3. Untied NMDS analysis, which understates the low proximity between unconnected structures (§4*a*), yields a solution in which dorsal and ventral visual areas are unequivocally segregated, and in which there is a hierarchical dimension extending between V1 and STPa/area 46.

4. NMDS configurations from visual system matrices that have been transformed by data conditioning methods that demonstrably abolish the effects of the data's sparsity (§6*b*,*c*) possess all the gross organizational features derived from the direct NMDS analysis (§6*d*).

(*b*) *Evidence on the reliability of NMDS analysis of this type of data*

5. The probability that the SSTRESS values for any of the NMDS analyses fall into the distribution of SSTRESS values for randomized data is vanishingly small (§4*a*; Young *et al.* 1994). Following the traditional reasoning in this area (Stenson & Knoll 1969), the probability that these NMDS analyses do not represent systematic aspects of data structure in the input matrix is correspondingly small.

6. The recovery of known parameters from untransformed test data that were as similar as possible to the visual system data by tied NMDS was characterized by approximately 96% variance-explained (§6*b*).

7. The recovery of known parameters from conditioned test data that were as similar as possible to the visual system data by NMDS was characterized by approximately 96% variance-explained (§6*b*).

(*c*) *Evidence from non-NMDS methods*

8. The quantitative distribution of connection strength specifies the manifolds that cortical neural systems can occupy in a Euclidean space, and the NMDS solutions lie in these shapes (§2*c* and §4).

9. The quantitative distribution of connection strength specifically prohibits cortical neural systems from taking the form of uniformly distributed disk structures (§2*c*).

10. χ^2 analysis of the connections between dorsal and ventral visual areas show that these groupings are much more connected within each grouping than between them (see §5*a*).

11. Similar analysis shows that striate and immediately prestriate areas and the high-order areas in the frontal and temporal lobes are much more connected within each grouping than between them.

12. χ^2 analysis of possible connections that have been confirmed absent between dorsal and ventral visual areas show that these groupings are much more disconnected between each grouping than within them.

13. Similar analysis shows that striate and immediately prestriate areas and high-order areas in the frontal and temporal lobes are much more disconnected between each grouping than within them.

14. Results from seriation analysis show that optimal length orderings for the visual system all place dorsal stream and ventral stream areas maximally far apart, separated by earlier visual areas at the one side, and by high-order visual areas at the other (see §5*c*).

15. NMDS and seriation, two independent analyses thus concur on the gross organizational features of the system. Statistical comparison of the arrangements of visual areas in the optimal length orderings with that in the NMDS analysis yields correlations that are approximately 0.9 (every $p < 0.000001$). The concurrence of the results of NMDS and seriation is thus not to be expected by chance.

16. Although the visual system matrix is certainly serially ordered, tour length is too blunt an analytical tool to tell disk structures apart from annuli (see §5*c*).

17. Hierarchical analysis of laminar origin and termination data constructs a largely self-consistent unidimensional hierarchy, in which the relative hierarchical level of particular structures can be made apparent (Felleman & Van Essen 1991).

18. Statistical comparison of the ordering of stations that emerges from hierarchical analysis shows that this ordering is statistically significantly related to the positions taken by the stations in the NMDS structure (p

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

< 0.000001). Two independent analyses of different types of anatomical data thus concur that: (i) the visual system is hierarchically organized; and (ii) about the approximate ordering of stations in the hierarchy, and do so to a degree that is not to be expected by chance.

19. Analysis of the behavioural effects of cortical lesions indicates that the visual system is dichotomized into a dorsal and a ventral pathway (Ungerleider & Mishkin 1982).

20. The organizational principles of the visual system that emerge from analysis of the behavioural effects of lesions can be captured in a simple numerical model. Statistical comparison of the ordering of cortical areas into the dorsal, ventral and shared groups in this model shows that this ordering is statistically significantly related to the positions taken by the areas in the NMDS structure ($p < 0.000001$). Two independent analyses of different types of anatomical data thus concur that: (i) the visual system is dichotomized; and (ii) about the classification of particular stations in the dichotomized system, and do so to a degree that is not to be expected by chance.

8. DISCUSSION

We have presented a detailed treatment of the application of NMDS to neuroanatomical connection data, and have explored the degree to which the results may be reliable. In the preceding section we enumerated the principal explananda in this area. We now discuss frameworks for explaining the results. We believe that all these results can be explained in a very economic manner. We suppose that NMDS, and χ^2 analysis of the connections and non-connections, seriation, hierarchical analysis and the analysis of the behavioural effects of cortical lesions have faithfully extracted underlying aspects of the structure from their particular types of data. These different methods variously and quantitatively agree that the system is divided into distinct streams of processing, each of which is hierarchically organized and which reconverge at multiple sites in the frontal and temporal lobes.

We note that alternative explanatory frameworks (Martin 1992; Goodhill *et al*. 1994; Simmen *et al*. 1994) fail to explain the majority of these results. It has been argued, for example, that parallel processing 'can reasonably be claimed in only one instance – in the topographic maps that repeat themselves again and again throughout the visual pathways' on the basis that the M and P pathways mix in V1 and V4, and that there is a non-striate pathway into MT (Martin 1992). This model would require that the dichotomization evident in results 1, 2, 3, 4, 10, 12, 14 and 19 (see §7) arose spuriously, and would have difficulty explaining results 5, 6, 7 and 8 which suggest that the NMDS configurations did not arise spuriously. In addition, this model does not explain how four independent methods concur that there is parallel processing, specifically within two gross subsystems of visual cortex, as in results 10, 12, 14, 15, 18 and 20. NMDS, χ^2, seriation and analysis of the behavioural effects of cortical lesions all agree that the system is dichotomized. Chance is an unlikely explanation of the agreement between the NMDS, seriation and lesion results (results 15, 18 and 20). It is vanishingly unlikely that artefact caused the agreement between these different methods, since the same artefact would have had to have afflicted such radically different analytical methods. This model appears, then, to require some further explication if it is to account for the facts.

Another alternative framework has recently been articulated (Goodhill *et al*. 1994; Simmen *et al*. 1994). In this model, it is suggested that 'the visual areas have an underlying abstract connectivity structure in which they are positioned throughout a circular region, reflecting little organization into streams' (Goodhill *et al*. 1994). This model is specifically excluded by the constraints of the quantitative distribution of connection strength (result 9), and is refuted by results 10 and 12. The model cannot then be true. Results 1, 2, 3 and 4, must have arisen spuriously in this model, despite the support of results 5, 6, 7 and 8. Results 14, 15, 19 and 20 are very difficult to explain within this framework, and as far as we know the authors of this model have made no attempt to explain the quantitative concordance in the results of all the independent methods.

In support of their views, however, Goodhill *et al*. (1994) have endeavoured to show that NMDS would artefactually find annular structure even if the true data structure were disk-like. To do this they binarized (Simmen *et al*. 1994) and quantized (Goodhill *et al*. 1994) the metric distances between points distributed within a disk. They showed that with 2s, 1s and 0s in the same proportion as the visual system data, NMDS reconstructed an unmistakeably annular configuration. In this way they hoped to specify a mechanism by which the divergence of the visual system into the distinct dorsal and ventral pathways evident in the NMDS structure (result 1) could come about artefactually: this organization could be explained by an annular bias in NMDS spuriously segregating the visual areas to different sides of the NMDS output space (Goodhill *et al*. 1994; Simmen *et al*. 1994).

We note first that disk structures cannot be specified by cortical connection data because of the constraints embodied by the quantitative distribution of connection strength (result 9 and §2*c*). Even taken at face value therefore, the finding of annularity from quantized disk data is irrelevant to the interpretation of the NMDS solution for the visual system, since disk data structure cannot arise in cortical connection data. We note secondly that this annular effect would, in any case, not explain result 2, that even when the visual areas are placed at the same side of the output space, the gross division of the system is still apparent. Neither would an annular bias explain how the system is still clearly divided into a dorsal and ventral stream when the annularity that arises from sparsity is abolished (result 4, §6*c*,*d*).

We note thirdly that the annular bias does not in fact arise from NMDS itself, as suggested by Goodhill *et al*. (1994), but from their quantization procedure and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

their misapplication of the tied approach in this case. Their quantization process transforms all distances above an arbitrary value into zeros, distances below this value but above another arbitrary quantization threshold into ones, and distances below this threshold into twos. The zeros in the quantized matrices correspond to a range of distances above an arbitrary threshold. In the case of this type of data, the comparison of the configuration derived from analysing the ternary data with the metric structure shows the degree to which NMDS can reconstruct the range of distances corresponding to each zero (or one or two). The appropriate NMDS approach in this circumstance is to allow tied similarities to untie. There is no special property of any of the zero entries – unlike connection data – that specifies that they should all be the same length, indeed the reverse is plainly the case. However, Goodhill *et al.* (1994) used the tied approach with these data. In tied NMDS, the algorithm finds a solution in which all the distances signalled by the same proximity value are as similar to each other as possible. Because zero proximitites signal the larger distances, they are emphasized in the optimization over the smaller distances implied by the non-zero proximities. Distances in the solution corresponding to non-zero proximities thus become more variable than distances corresponding to zero proximities. The result is that the solution is one in which the frequency distribution of distances possesses a peak for long distances, and a smoother and lower distribution of shorter distances. We demonstrated in §2*c* that rings or horseshoes have proximity distributions with this shape. This process for turning disks into rings requires at least two properties that connection data do not possess. First, the initial data structure must be that of a disk. Cortical connection data cannot possess this data structure (§2*a*,*c*). Second, the zero proximities in the connection matrix must correspond to a range of distances. Unless one absent connection can reasonably be said to be more or less absent than another absent connection, there is no range of proximities underlying the zero proximities in connection data. Hence, the particular suggestion from these colleagues (Goodhill *et al.* 1994; Simmen *et al.* 1994) is not a candidate mechanism for the etiology for the annularity or curvature of NMDS structures for cortical neural systems, and they present no evidence that these NMDS structures are corrupted by artefact. Their test data are fundamentally unlike real connection data. Their comments reflect, we believe, an understandable concern about the ubiquitous curvature of NMDS analyses of cortical connection data. It is not, however, a matter of opinion that this curvature arises as an inevitable consequence of *bona fide* aspects of data structure, and that the particular problem exhibited by Simmen *et al.* (1994) and Goodhill *et al.* (1994) does not apply to analysis of connection data.

These considerations, and those summarized in §7, reflect the fact that the evaluation of the gross organization of central neural systems has entered the area of quantitative biology. The success of a particular proposition now turns on its ability to account for quantitative facts, rather than on its intuitive appeal, or the tradition from which it arises. The three propositions that the visual system is hierarchically organized, divided into gross streams and that it provides opportunities for processed visual signals to reconverge, account for results from five independent methods of analysis of three different types of neurobiological data that bear on the organization of the system. We have also shown by simulation that one of these independent methods of analysis, NMDS, reliably recovers the patterns underlying test data that were as similar as possible to connection data. The mutual corroboration between multiple independent methods and our demonstration of the reliability of NMDS in this application suggest that these three organizing principles are presently as well supported as any set of conclusions in quantitative biology. We believe that these results imply that the features of the primate cortical visual system are approximated closely and clearly in the structures derived from NMDS analysis of the system's connection data, and we look forward to further experimental tests of the insights on this and other cortical systems that emerge from considering these structures.

Supported by the Wellcome Trust, SERC, MRC, the McDonnell–Pew Centre for Cognitive Neuroscience, and the Royal Society. We are grateful to Dr R. J. Baddeley for his comments on an earlier draft of the manuscript.

REFERENCES

Aggleton, J.P., Burton, M.J. & Passingham R.E. 1980 Cortical and subcrtical afferents to the amygdala of the rhesus monkey (*Macaca mulatta*). *Brain Res.* **190**, 347–368.

Amaral, D.G. & Price, J.L. 1984 Amygdalo-cortical projections in the monkey (*Macaca fascicularis*) *J. comp. Neurol.* **230**, 465–496.

Barbas, H. 1988 Architectonic and connectional organization of ventral and dorsal prefrontal areas in the rhesus monkey. *Epilepsia* **29**, 209–210.

Bowman, E.M. & Olson, C.R. 1988 Visual and auditory association areas of the cat's posterior ectosylvian gyrus: cortical afferents. *J. comp. Neurol.* **272**, 30–42.

Cavada, C. & Goldman-Rakic, P.S. 1989 Posterior parietal cortex in Rhesus monkey: I. Parcellation of areas based on distinctive limbic and sensory corticocortical connections. *J. comp. Neurol.* **287**, 393–421.

Cherniak, C. 1994 Component placement optimization in the brain. *J. Neurosci.* **14**, 2418–2427.

Colby, C.L. & Duhamel, J.-R. 1991 Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey. *Neuropsychologia* **29**, 517–537.

Coombs, C.H. 1964 *A theory of data*. New York: Wiley.

Cowey, A. 1979 Cortical maps and visual perception. *Q. J. exp. Psychol.* **31**, 1–17.

Dejerine 1895 *Anatomie des centres nerveux*. Paris.

Dempsey, P. & Baumhoff, M. 1963 The statistical use of artifact distributions to establish chronological sequence. *Am. Antiquity* **28**, 496–509.

De Yoe, E.A. & Van Essen, D.C. 1985 Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature, Lond.* **317**, 58–61.

Edgington, E.S. 1980 *Randomization tests*. New York: Marcel Dekker.

Engel, A.K., Konig, P., Kreiter, A.K., Schillen, T.B. & Singer W. 1991 Temporal coding in the visual cortex:

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

new vistas on integration in the nervous system. *Trends Neurosci.* **15**, 218–226.

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

Gerfen, C.R. & Sawchenko P.E. 1984 An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons and terminals – immunohistochemical localization of an axonally transported plant lectin, Phaseolus-vulgaris leukoagglutinin (PHA-L). *Brain Res.* **290**, 219–238.

Goodale, M.A. & Milner, D.A. 1993 Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25.

Goodhill, G.J., Simmen, M.W. & Willshaw D.J. 1994 An evaluation of the use of multidimensional scaling for understanding brain connectivity. Research Paper EUCCS/RP-63, Centre for Cognitive Science, University of Edinburgh.

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

Gower, J.C. & Legendre P. 1986 Metric and Euclidean properties of dissimilarity coefficients. *J. Class.* **3**, 5–48.

Hole, F. & Shaw, M. 1967 Computer analysis of chronological seriation, vol. 53. Rice University Studies.

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

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. stat. Psychol.* **27**, 133–153.

Katz, L.C. & Iarovici, D.M. 1990 Green fluorescent latex microspheres – a new retrograde tracer. *Neuroscience* **34**, 511–520.

Kendall, D.G. 1971 Seriation from abundance matrices. In *Mathematics in the archaeological and historical sciences* (ed. F. R. Hodson *et al.*), 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.

King, M.A., Louis, P.M., Hunter, B.E. & Walker, D.W. 1989 Biocytin - a versatile anterograde neuroanatomical tract–tracing alternative. *Brain Res.* **497**, 361–367.

Kruskal, J.B. 1964a Multidimensional scalingby optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* **29**, 1–27.

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

Kuzara, R.S., Mead, R.G. & Dixon, K.A. 1966 Seriation of anythropological data: a computer program for matrix ordering. *Am. Anthropol.* **68**, 1442–1455.

Laporte, G. & Taillefer S. 1987 An efficient interchange procedure for the archaeological seriation problem. *J. archaeol. Sci.* **14**, 283–289.

de Leeuw, J., Young, F.W. & Takane Y. 1976 Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika* **41**, 471–503.

Lefkovitch, L.P. 1991 Vector dissimilarity and clustering. *Math. Biosci.* **104**, 39–48.

Martin, K.A.C. 1992 Parallel pathways converge. *Curr. Biol.* **2**, 555–557.

Maunsell, J.H. & Van Essen D.C. 1983 The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* **3**, 2563–2586.

Merigan, W.H. & Maunsell J.H.R. 1993 How parallel are the primate visual pathways? *A. Rev. Neurosci.* **10**, 363–401.

Mitchison, G. 1991 Neuronal branching patterns and the economy of cortical wiring. *Proc. R. Soc. Lond.* B **245**, 151–158.

Mitchison, G.J. & Durbin R. 1986 Optimal numberings of an NxN array. *SIAM Jl Algor. Discrete Meth.* **7**, 571–582.

Musil, S.Y. & Olson C.R. 1988*a* Organization of the cortical and subcortical projections to the anterior cingulate cortex in the cat. *J. comp. Neurol.* **272**, 203–218.

Musil, S.Y. & Olson C.R. 1988*b* Organization of the cortical and subcortical projections to the medial prefrontal cortex in the cat. *J. comp. Neurol.* **272**, 219–241.

Musil, S.Y. & Olson, C.R. 1991 Cortical areas in the medial frontal lobe of the cat delineated by quatitative analysis of thalamic afferents. *J. comp. Neurol.* **308**, 457–466.

Nicolelis, M.A.L., Tinone, G., Sameshima, K., Timo-Iaria, C., Hong, Y.C. & Van de Bilt, M.T. 1990 Connection, a microcomputer program for storing and analyzing structural properties of neural circuits. *Comput. Biomed. Res.* **23**, 64–81.

Olson, C.R. & Jeffers, I.M. 1987 Organization of cortical and subcortical projections to area 6m in the cat. *J. comp. Neurol.* **266**, 73–94.

Olson, C.R. & Lawler, K. 1987 Cortical and subcortical afferent connections of a posterior division of feline area 7 (area 7p). *J. comp. Neurol.* **259**, 13–30.

Olson, C.R. & Musil, S.Y. 1992 Topographic organization of cortical and subcortical projections to posterior cingulate cortex in the cat: evidence for somatic, ocular, and complex subregions. *J. comp. Neurol.* **324**, 237–260.

Perrett, D.I. & Oram, M.W. 1995 Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque: I. Evidence for information binding and classification of response types. (In the press.)

Pribram, K.H. & MacLean, P.D. 1953 Neuronographic analysis of medial and basal cerebral cortex. II. Monkey. *J. Neurophysiol.* **16**, 324–340.

Rockland, K.S. & Pandya, D.N. 1979 Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res.* **179**, 3–20.

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

Scannell, J.W., Young, M.P. & Blakemore, C. 1995 Analysis of connectivity in the cat cerebral cortex. *J. Neurosci.* **15**, 1463–1483.

Schonemann, P. & Carroll, R.M. 1970 Fitting one matrix to another under choice of a similarity transformation and a rigid motion. *Psychometrika* **35**, 245–255.

Shepard, R.N. 1958 Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychol. Rev.* **65**, 242–256.

Shepard, R.N. 1962 Multidimensional scaling with an unknown distance function. *Psychometrika* **27**, 125–140.

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

Shepard, R.N. & Carroll, J.D. 1969 PARAMAP. In *Multivariate analysis* (ed. P. R. Krishnaia), 561 pp. New York: Academic Press.

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

Stefan, A. 1971 Application of mathematical methods to epigraphy. In: *Mathematics in the archaeological and historical*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sciences (ed. F.R. Hodson *et al.*), pp. 267–275. Edinburgh University Press.

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

Takane, Y., Young, F.W. & de Leeuw, J. 1977 Non-metric individual differences multidimensional scaling: an alternative 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.

Turner, B.H., Mishkin, M. & Knapp, M. 1980 Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. *J. comp. Neurol.* 24 **191**, 515–543.

Ungerleider, L.G. & Desimone, R. 1986 Cortical connections of visual area MT in the macaque. *J. comp. Neurol.* **248**, 190–222.

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

Van Essen, D.C. & De Yoe, E.A. 1995 In *The cognitive neurosciences* (ed. M. S. Gazzaniga). Boston, Massachussetts: MIT Press.

Vieussens 1685 *Neurographia.*

Vidalsanz, M., Villegasperez, M.P., Bray, G.M. & Aguayo, A.J. 1988 *Expl Neurol.* **102**, 92–101.

Whitlock, D.G. & Nauta, W.J.H. 1956 Subcortical projections from temporal neocortex in Macaca mulatta. *J. comp. Neurol.* **106**, 183–212.

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

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

Young, M.P. 1990 Exploratory across-stimulus studies in event-related potentials. Ph.D. thesis, St Andrews University.

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. 1995 Analysis of connectivity: systems in the cerebral cortex. *Rev. Neurosci.* **5**, 227–250.

Received 3 *October* 1994; *accepted* 16 *December* 1994

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES