Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac)

**Klaas E. Stephan^{1,2}, Lars Kamper¹, Ahmet Bozkurt¹, Gully A. P. C. Burns³,
Malcolm P. Young² and Rolf Kötter^{1,4*}**

¹*Computational Systems Neuroscience Group, C. and O. Vogt Brain Research Institute, Heinrich Heine University Düsseldorf,
Moorenstrasse 5, 40225 Düsseldorf, Germany*

²*Neural Systems Group, Department of Psychology, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK*

³*Knowledge Mechanics Research Group, Department of Neurobiology, University of Southern California, Los Angeles,
CA 90089-2520, USA*

⁴*Institute of Morphological Endocrinology and Histochemistry, Heinrich Heine University Düsseldorf, Moorenstrasse 5,
40225 Düsseldorf, Germany*

CONTENTS

1. Introduction	1160
2. Methods	1161
(a) Conceptual rationale	1161
(b) Data representation	1162
(i) Main classes of data	1162
(ii) Mapping data: brain sites, brain maps and inter-map relationships	1162
(iii) Experiment-related data	1165
(iv) PDCs: coding the precision of descriptions	1169
(v) Relational structure of CoCoMac	1171
(vi) Standardized principles of data collation and representation	1172
(c) Data analysis	1172
(i) General concepts of data analysis and integration in CoCoMac	1172
(ii) Preprocessing of primary data	1173
(iii) ORT: transformation of integrated projection data to user-chosen target map	1176
(iv) Post-processing of transformation results	1177
(v) User-defined parameters of data analysis	1177
(d) Technical implementation	1177
3. Results	1177
(a) Current contents	1177
(b) Analysing prefrontal connectivity: an example of the practical use of CoCoMac	1177
4. Discussion	1180
Appendix A	1182
Appendix B	1183
Appendix C	1183
References	1184

*Author for correspondence (rk@hirn.uni-duesseldorf.de).

The need to integrate massively increasing amounts of data on the mammalian brain has driven several ambitious neuroscientific database projects that were started during the last decade. Databasing the brain's anatomical connectivity as delivered by tracing studies is of particular importance as these data characterize fundamental structural constraints of the complex and poorly understood functional interactions between the components of real neural systems. Previous connectivity databases have been crucial for analysing anatomical brain circuitry in various species and have opened exciting new ways to interpret functional data, both from electrophysiological and from functional imaging studies. The eventual impact and success of connectivity databases, however, will require the resolution of several methodological problems that currently limit their use. These problems comprise four main points: (i) objective representation of coordinate-free, parcellation-based data, (ii) assessment of the reliability and precision of individual data, especially in the presence of contradictory reports, (iii) data mining and integration of large sets of partially redundant and contradictory data, and (iv) automatic and reproducible transformation of data between incongruent brain maps.

Here, we present the specific implementation of the 'collation of connectivity data on the macaque brain' (CoCoMac) database (<http://www.cocomac.org>). The design of this database addresses the methodological challenges listed above, and focuses on experimental and computational neuroscientists' needs to flexibly analyse and process the large amount of published experimental data from tracing studies. In this article, we explain step-by-step the conceptual rationale and methodology of CoCoMac and demonstrate its practical use by an analysis of connectivity in the prefrontal cortex.

Keywords: macaque; structure–function relationships; prefrontal cortex; neural networks; analysis; relational database

1. INTRODUCTION

Major efforts have been started to provide the neuroscientific community with powerful databases of various data modalities. In analogy to the indispensable role of bioinformatics for molecular biology, neuroinformatics has gained considerable importance for the investigation of complex structure–function relationships in the nervous system (Kötter 1999; Young & Scannell 2000). Some of the most notable database projects in neuroinformatics have concerned data on anatomical connectivity, in particular association fibres that connect discriminable brain structures over long distances (Burns 1997, 1999; Felleman & Van Essen 1991; Scannell *et al.* 1995, 1999; Young 1993). The considerable efforts of constructing these databases are warranted by the key role that structural connectivity data play for analyses of brain organization. Anatomical connections represent structural constraints of the complex and poorly understood functional dynamics of neural systems. Any explanation of brain function will be unsatisfactory unless it integrates the structural network in which the functional processes are undertaken. Concerning the human brain, there still is a complete lack of detailed connectivity data due to methodical and ethical restrictions. On the other hand, thousands of anatomical tracing studies have been performed during the last decades in different mammalian species. The ultimate aim of these studies is to obtain a complete map of each brain's communication network that is constituted by its fibre connections. The current state of this enormous project is indeed available to the neuroscientific community—but only in the form of fragments distributed over thousands of individual reports in the literature, based on different species, different tracing methods, and different types of data representation. This is the paradigmatic situation where powerful databases are needed. The obstacles that impede the construction of such connectivity databases, however,

are considerably larger than in molecular biology. There are three main reasons for this.

First, descriptions of anatomical projections are complex multidimensional data. In addition to source and target structures, many additional properties of a particular projection are of interest, e.g. strength, laminar patterns and other spatial distributions of sending and receiving neurons, or quantitative data like absolute numbers of neurons. Unfortunately, no universally agreed standards have been established for how anatomical projections and their multiple properties should be represented in the literature or in databases. Genome sequence data, in comparison, are commonly represented by the linear sequence of base pairs, i.e. as a unidimensional sequence of a simple four-letter code.

Second, the data are not only complex, but also highly method-dependent. Representation of data would be simpler and more efficient if one could separate the method-related part from the method-independent characteristics. In pharmacology, as an analogy, receptors used to be classified with reference to the differential binding of a set of ligands. Therefore, classifications of receptors depended on the selectivities and relative affinities of the available ligands. When the sequences of receptor proteins became known, it was possible to classify receptors based on sequence similarities without reference to the method used for the sequencing. This improvement indicates that many neuroscientific data are not optimally parameterized or even lack discrete properties altogether.

Third, all available tracing studies report their findings on the basis of brain maps (as opposed to spatial coordinates). Due to a variety of different methods that are used for the parcellation of individual brain structures and because of the considerable observer-dependence of these methods, a large number of different parcellation schemes have been defined for most regions of any species' brain. Such competing maps are often incompatible, with respect

both to the boundaries of their delineated brain structures and to their nomenclature.

In spite of these profound difficulties, a first generation of connectivity databases had been constructed throughout the last decade: for the macaque monkey (Felleman & Van Essen 1991; Young 1993), cat (Scannell *et al.* 1995, 1999) and rat (Burns 1997). These databases have led to numerous successful investigations that have shed new light onto complex structure–function relationships in mammalian brains. Not only have they been crucial for sophisticated theoretical analyses of brain networks (Burns & Young 2000; Felleman & Van Essen 1991; Graepel & Obermayer 1999; Hilgetag *et al.* 1996, 2000*a,b*; Jouve *et al.* 1998; Kötter & Sommer 2000; Scannell 1997; Scannell *et al.* 1995, 1999; Sporns *et al.* 2000; Young 1992, 1993; Young *et al.* 1995, 2000) but they also stimulated numerous experimental investigations of predictions that arose from them (Barone *et al.* 2000; MacNeil *et al.* 1997; Merabet *et al.* 1998; Scannell *et al.* 1996).

Notwithstanding these successes, some inherent methodological limitations of the first-generation databases impeded further progress. Most importantly, the parcellation problem had not been solved convincingly. The first-generation databases approached the problem of the many divergent parcellation schemes by defining a ‘reference map’ *a priori* before the process of data collation was started. When entering published data into the database, the original findings were mapped to the reference map according to the judgement and individual criteria of the respective data collator. This approach made it difficult to identify and correct errors or apply reinterpretations of findings. Also, it restricted the output from the database to one fixed brain map, which did not allow flexible data extraction by the user. An additional problem was that summary representations were used predominantly: all primary reports that concerned a particular projection $X \rightarrow Y$ (as described by the reference map) were condensed into a single entry of the database, with the original data and the criteria of merging them being visible to the user only by further detailed comparison between the original literature and the database itself. The resulting mode of data representation therefore was not satisfactorily independent of the collator (for a more detailed discussion of these problems see Stephan *et al.* (2000*a*)).

In spite of these problems, the first generation of connectivity databases were pioneering attempts to introduce powerful database technologies to systems neuroscience. They have substantially contributed to irreversibly changing neuroscientific concepts of how to investigate complex brain systems. Also, they provided highly valuable experiences and stimulation for the development of current connectivity databases.

In this article, we describe in detail the methodological rationale of CoCoMac, a connectivity database of the macaque monkey. Originally conceived as an acronym of ‘cortical connectivity of the macaque brain’, CoCoMac now comprises subcortical connections as well. Thus, at present, the acronym is better understood as a ‘collation of connectivity data on the macaque brain’. We shall first introduce a list of general theoretical axioms for connectivity databases (§2*a*); while explaining the principles of

CoCoMac step by step, we shall demonstrate in what way its conceptual design reflects these axioms. In §2*b* and §2*c*, respectively, we shall separately illustrate the distinct processes of data representation and data analysis within CoCoMac. Finally, we shall outline its current contents (§3*a*) and provide a simple example of its practical use (§3*b*).

Generally, the aim of this article is not to deliver a complete technical description of CoCoMac, but rather to explain its general conceptual rationale. Mathematical definitions are only referred to where helpful for the general understanding of our implementation. The reader who is interested in details of the implementation is referred to our previous publications on objective relational transformation (ORT) (Stephan & Kötter 1998, 1999; Stephan *et al.* 2000*a*) and forthcoming publications that cover the technical aspects of CoCoMac in mathematical detail.

2. METHODS

(a) Conceptual rationale

Methodologically, building a database that contains experimental data from the literature is equivalent to ‘observing’ and interpreting the results of experiments conducted by various persons and representing these observations and interpretations on a certain scale. Therefore, collating data into databases should satisfy the same three criteria that are generally applied to both tests and measurements in the natural sciences: (i) objectivity (i.e. observer-independence or interobserver reproducibility), (ii) reliability (i.e. intra-observer reproducibility), and (iii) validity (Bortz & Döring 1995; Guilford 1954; Lienert & Raatz 1994). Transferred to the context of databasing, objectivity means that different observers should ideally yield equivalent (or, at least, sufficiently similar) representations of data when dealing with the same primary data source in the literature. Reliability means that repeated collations of the same set of original data by the same observer should ideally yield equivalent representations within the database. Validity concerns the question of whether the data representations within the database reflect the properties of the original data in a meaningful way.

When we started constructing CoCoMac, we initially had two main goals: first, the architecture of CoCoMac should strive to approximate the ideal of objective, reliable and valid data representations. Second, CoCoMac should offer high flexibility to both users and developers. Aiming at these two goals and learning from the strengths and weaknesses of the first-generation connectivity databases (see §1), we defined seven axioms that guided the development of CoCoMac:

- (i) Strict separation of data representation and data analysis. Data representation and data analysis should represent two entirely different levels of database functionality. Data representation should not be ‘contaminated’ by any implicit analyses or data conversions (e.g. merging findings across several original reports into one summary entry). Instead, different tools of analysis may be applied once the process of data representation is completed, thus allowing for different perspectives on the original data.
- (ii) 1:1-representation of original data. Each original statement should be entered into the database, including all redundant and contradictory findings within and across different articles. Furthermore, no transformations of the notation of

brain data to a predefined 'reference map' should be performed; instead, each original finding should be described by its original nomenclature and parcellation scheme. All entries in the database should be linked to their source in the literature by exact references (citations, page and figure numbers).

- (iii) Standardized collation of data. The process of data collation itself should be strictly regulated to ensure that representations of the same set of data in CoCoMac by different observers are as equivalent as possible. Data collation should be split into a series of distinct steps. Clear rules should define the order in which these steps are executed and how they are performed exactly.
- (iv) Transparent data analysis. Analyses and transformations of the data should be performed algorithmically. The concepts of these algorithms and the impact of various parameters on the analyses should be transparent and published (Stephan *et al.* 2000a; <http://www.cocomac.org>). Furthermore, the user should be able to look up the chain of analytical steps that led to a particular result.
- (v) User-defined analysis. The user should be able to define all free parameters of the analyses according to individual preferences. This also includes the choice of a target map for the transformation of data between different parcellation schemes.
- (vi) Flexible extensibility. The relational design of CoCoMac should be flexible enough to make future extensions possible without great efforts. For example, one should be able easily to add new objects of data representation (e.g. on the single-neuron level) or further data modalities in CoCoMac that complement connectivity data (e.g. distributions of transmitter receptors).
- (vii) Applicability to existing data. The huge reservoir of experimental data from anatomical tract tracing studies is characterized by a startling variability in descriptive precision, nomenclature, and parcellation schemes across different articles. The conceptual design of CoCoMac should be formally strict to ensure that all original data are treated in the same way regardless of their presentation by the literature. Yet it also needs to account for this diversity in a way that the predefined procedures of data collation can be easily applied to the entire range of connectivity reports.

We shall now specify how the actual design of CoCoMac reflects these axioms. It contains three main parts, describing separately the general principles of data representation (§2b), data analysis (§2c) and technical implementation (§2d).

(b) *Data representation*

(i) *Main classes of data*

The data contained by CoCoMac can be grouped into three main classes: bibliographical data, mapping data and experiment-related data. To provide a brief survey of these three groups, we list the main tables of CoCoMac in table 1. The group of mapping data contains some fundamental objects (brain sites and inter-map relationships) of CoCoMac that provide the basis for a representation of experimental results by their original nomenclature and parcellation scheme. Therefore, we first discuss tables and objects for mapping data before we address the representation of experimental data.

(ii) *Mapping data: brain sites, brain maps and inter-map relationships*

Available tracing studies do not describe their results by coordinates in Euclidean three-dimensional (3D) space, but instead refer to descriptive entities which are defined by microstructural (e.g. cytoarchitectonics, receptor or enzyme distributions) or functional (e.g. electrophysiological) properties. Classical anatomical nomenclature calls these entities 'areas' or 'fields' in the cortex, and 'nuclei' in the subcortical structures. Within CoCoMac, these entities are generally referred to as 'brain sites' (or, if addressed as technical objects, as *BrainSite* objects). Any given brain site is addressed by a string $\langle A_1A_2 \dots A_nY-X \rangle$ in CoCoMac, with $A_1A_2 \dots A_n$ being the initials of the authors of the article, Y the two-digit year of publication, and X the acronym by which these authors refer to the brain site (for example, medial area 11 as defined by Carmichael & Price (1994) and abbreviated by them as '11m' is referred to by CoCoMac as 'CP94-11m'). A set of such microstructurally functionally defined brain sites as presented by a particular article constitutes a 'parcellation scheme' or 'brain map' in CoCoMac. In analogy to a spatial description of brain data in Euclidean 3D space, describing brain data on the basis of parcellation schemes can be envisaged as representations in a non-Euclidean multidimensional space of microstructural–functional properties. For example, in an architectonic study on prefrontal cortex, Barbas & Pandya (1989) delineated areal boundaries on the basis of: (i) width of granular layer IV, (ii) density of pyramidal cells in layer III, (iii) density of pyramidal cells in layer V, and (iv) overall cell density in the supragranular layers. Together, these criteria define the four dimensions of a non-Euclidean feature space in which the delineated areas can be located.

Generally, data are only directly comparable if they are described within the same space—otherwise, a mapping function is needed to transform the data into one common space. In our context, comparing data that are described by means of different parcellation schemes (=different microstructural–functional feature spaces) requires us to map the brain sites from the various parcellation schemes involved into one common brain map. Unfortunately, due to a variety of applied methods and the high observer-dependency of many microstructural–functional criteria, there exist a large number of parcellation schemes that differ in nomenclature, number, location, size and boundaries of the delineated brain sites. This variability is especially high for the cortex, but is also regularly found for subcortical structures (e.g. thalamus, amygdala and brain stem). In this situation, comparisons of data described by two different parcellation schemes can be very difficult. Formally, such a comparison is equivalent to comparing two points from two different non-Euclidean spaces whose dimensions are characterized by the microstructural–functional criteria of the two parcellation schemes. If these criteria are different in number and/or quality between the two maps, the two feature spaces consequently also have different numbers and/or different types of dimensions. The question of how one can compare data based on two parcellation schemes can thus be rephrased as 'Is it possible to define a mapping function between these two spaces?' or, less formally, 'Is there a way to relate the brain sites of the two parcellations schemes to each other so that they may be compared in a meaningful way?' Since spatial coordinates of the brain sites are not provided, CoCoMac uses the concept of 'inter-map relations' (or, if addressed as technical objects, as *InterMapRelation* objects). These include all possible logical relationships that two brain sites A and B from different brain maps A' and B' can have and are

Table 1. Survey of main tables for data representation in CoCoMac. Tables can be classified as containing either bibliographical, mapping or experiment-related data.

bibliographical data	
<i>Literature</i>	general information on tracing reports and parcellation schemes in the literature
<i>Literature_JournalArticles</i>	information on journal articles
<i>Literature_Books</i>	information on books
<i>Literature_BookChapters</i>	information on book chapters
<i>Literature_Authors</i>	information on authors
mapping data	
<i>BrainMaps</i>	list of all brain maps
<i>BrainMaps_Delineation</i>	information on delineated brain sites of a particular brain map
<i>BrainMaps_UnspecificAdoption</i>	information on unspecifically adopted brain sites of a particular brain map
<i>BrainMaps_Methods</i>	information on experimental methods and criteria used to delineate brain sites
<i>BrainMaps_BrainSites</i>	list of all brain sites
<i>BrainMaps_BrainSiteAcronyms</i>	acronyms of brain sites and their explanations
<i>InterMapRelations</i>	logical relations between brain sites of different maps
<i>InterMapRelations_References</i>	references for statements on inter-map relationships in the literature
experiment-related data	
<i>Methods</i>	methodical information on the individual experiments
<i>Methods_Animals</i>	information on animals (macaque subspecies, gender, age, weight)
<i>Methods_RadioactiveTracers</i>	additional methodical information on experiments with radioactive tracers
<i>Injections</i>	information on individual injections
<i>Injections_AffectedSites</i>	information on injections that affected more than one brain site
<i>Injections_Laminae</i>	information on which laminae were affected by a particular injection
<i>LabelledSites_Descriptions</i>	general information about label resulting from a particular injection
<i>LabelledSites_Data</i>	specific information about the brain sites labelled by a particular injection
<i>LabelledSites_Data_Laminae</i>	information on which laminae were labelled within a particular brain site
<i>LabelledSites_Data_QD</i>	several tables with quantitative data on labelled brain sites

presented within CoCoMac by the relation codes (RCs) (see Stephan *et al.* 2000a):

- (i) RC(A, B) = I: A and B have identical boundaries.
- (ii) RC(A, B) = S: A is a subarea of B, i.e. A is contained by B.
- (iii) RC(A, B) = L: A is larger than B, i.e. A contains B.
- (iv) RC(A, B) = O: A and B overlap, i.e. A includes some parts of the brain that are not included by B and vice versa.
- (v) RC(A, B) = D: A and B are disjoint, i.e. A and B are not coextensive in any way.

To explain the notion of inter-map relationships in more depth, we take an example from some brain maps of premotor cortex. For the two studies of Brodmann (1909) and Von Bonin & Bailey (1947), the experimental method (i.e. Nissl staining) and the criteria used for delineation (i.e. distinctive cytoarchitectonic properties determined by visual inspection) are very similar, but of relatively low observer-independence. In spite of the resulting obvious differences between these two maps, one can still compare these parcellations relatively easily because the dimensions of the two microstructural spaces are of similar quality. One finds, for example, that the general cytoarchitectonic features described by Brodmann for his area 6 (e.g. agranular cortex without Betz cells in layer V) are still valid for each of the premotor areas FB, FBA, and FCBm recognized by Von Bonin & Bailey, but that the latter authors used additional criteria to distinguish these subdivisions within Brodmann's area 6. Areas FB, FBA, FCBm of Von Bonin & Bailey can thus be defined as subareas of Brodmann's area 6. A third study by Vogt & Vogt (1919) also follows a classical architectonic approach, using visually determined changes in myeloarchitecture to detect areal borders. If this parcellation is compared with either Brodmann's (1909) or Von Bonin & Bailey's (1947),

how can the microstructural spaces be related to each other, as their dimensions are of an entirely different quality? In other words, how can we determine whether an areal border that has been determined by purely cytoarchitectonic criteria is matched by a border determined by purely myeloarchitectonic criteria? This direct comparison is possible if there is a way of establishing a clear match between particular myeloarchitectonic and particular cytoarchitectonic features. Studies combining both criteria (e.g. Barbas & Pandya 1987) can indeed serve as a connecting link that allows such a comparison. Beyond mere architectonic criteria, similar problems of dimensional mismatch between the feature spaces of different parcellations are found if any of the structural studies above is compared with other studies of premotor cortex that use electrophysiological criteria to distinguish areas on the basis of different response properties (e.g. Mitz & Wise 1987).

It seems that in such cases where the microstructural–functional feature spaces of two parcellations are not directly comparable, additional information is needed which must arise from the parcellations themselves, not from their underlying criteria. Commonly, this is topographical information that results from a comparison of the two parcellations with respect to relative size and position of areas and relative location to landmarks like sulci. This process suffers from considerable observer-dependence and is additionally confounded by the high interindividual differences between brains. Nevertheless, rough topographical information is always available for two given parcellation schemes and may be the only way to perform any comparison at all for parcellations with no direct comparability of their microstructural–functional feature spaces. Both approaches—comparisons of different parcellation schemes by microstructural–functional and topographical criteria—are

frequently found in the literature. Unfortunately, very few authors explicitly reveal the basis of their statements on relationships between different brain maps. Often, a combination of both microstructural–functional and topographical comparisons seems to underlie such statements.

BrainSites and *InterMapRelations* as defined above are the most fundamental objects for data representation in CoCoMac. In order to explain how exactly they are used within CoCoMac, we begin with a simple example: after specifically adopting the parcellation schemes of Walker (1940) and Pandya & Seltzer (1982) for prefrontal and parietal areas, respectively, McGuire *et al.* (1991) report an anterogradely traced projection from Walker's area 46 (W40-46) to Pandya & Seltzer's area PG (PS82-PG). This reported projection is represented in CoCoMac by stating that the injection was placed in W40-46 and that anterogradely labelled terminals were found in PS82-PG. All further data related to this projection (e.g. strength, laminar patterns of neurons or terminals) are also described by reference to these two brain sites W40-46 and PS82-PG. In this way, we achieve a representation of the original data that is free from personal interpretations but exactly follows the nomenclature and parcellation scheme used by the original report.

Second, brain sites are also used to represent statements about inter-map areal relationships. For example, Walker's (1940) area 46 has been stated to lie within Brodmann's (1909) area 9 (see Von Bonin & Bailey 1947, p. 66), and Pandya & Seltzer's (1982) area PG is considered to be a subpart of Brodmann's area 7 (see Preuss & Goldman-Rakic 1991, p. 491 and fig. 9C). Thus, these relationships can be represented in CoCoMac as $RC(W40-46, B09-9) = S$ and $RC(PS82-PG, B09-7) = S$. Representing brain sites and inter-map relationships in this way preserves objectivity and reliability of the data, yet also allows later transformation of the original finding $W40-46 \rightarrow PS82-PG$ to the map of Brodmann ($B09-9 \rightarrow B09-7$) or to any other map whose relationships to W40-46 and PS82-PG are known in CoCoMac. This conversion is done algorithmically by means of ORT (see Stephan & Kötter 1998, 1999; Stephan *et al.* 2000a). ORT not only guarantees a transparent, reproducible, and mathematically defined transformation of connectivity data, but it is also useful to deduce unknown inter-map relationships from already known ones (see § 2c(ii) and § 2c(iii)).

Having introduced *BrainSites* and *InterMapRelations* as the basic objects for data representation, we need to take a closer look at how brain sites are described by articles and when we actually need to create a representation of a brain site in CoCoMac. We distinguish the following general ways of defining brain sites in a given article:

- (i) **Delineation.** The article explicitly defines locations and boundaries of the brain sites it refers to. The parcellation may either be a new one or confirm existing ones—in either case, the delineated brain sites may be adopted or referred to by other articles. The definitions of the brain sites are given by textual descriptions of the delineated areas (e.g. extent, location, architectonic properties of the brain site) and/or by figures (overviews or sections) on which the boundaries and/or architectonic properties of the brain sites are shown.
- (ii) **Unspecific adoption.** The article does not delineate brain sites itself, but adopts brain sites from one or several other parcellation schemes. The relationship between the brain sites and the parcellation schemes, however, is not unambiguously specified insofar as it is not stated what

parcellation scheme a particular brain site is taken from. This situation arises when a brain site is used without any reference to a brain map at all (e.g. Carmichael & Price (1995b) refer to areas 7a and 7b but provide neither a definition nor a reference). Alternatively, several parcellations with similar or identical names for brain sites are referred to simultaneously without exact specification of which definition applies to a particular brain site. For example, in a study of amygdalar connectivity, Aggleton (1986, p. 516) refers to several parcellation schemes that are only partially compatible: 'The nomenclature and cytoarchitectonics of the amygdaloid nuclei have been described elsewhere (Crosby & Humphrey 1941; Jimenez-Castellanos 1949; Aggleton & Mishkin 1984)'. Generally, parcellation schemes are often difficult to distinguish by mere nomenclature as they may contain similar and sometimes identical names, but may still differ in their definitions even for identically named areas. For example, Carmichael & Price (1994), Matelli *et al.* (1991) and Vogt *et al.* (1987) each describe an area 24c. In spite of the identical acronyms, these three definitions of area 24c are all different (see the statements of Carmichael & Price (1994, p. 389), and Matelli *et al.* (1991, p. 460), on the relationships between these definitions).

- (iii) **Specific adoption.** The article does not delineate brain sites itself, but adopts brain sites from one or various other parcellation schemes. In contrast to unspecific adoption ((ii) above), the source for the adopted brain sites is clearly specified. For example, McGuire *et al.* (1991, p. 394) state their sources of delineation precisely: 'Every cortical area that contained labelled terminals or cells was identified with thionin-stained sections, using the cytoarchitectural descriptions of Walker (1940; prefrontal cortex), Barbas & Pandya (1987; premotor cortex and SMA [supplementary motor area], Von Bonin & Bailey (1947; motor cortex), Powell & Mountcastle (1959; somatosensory cortex) ...'.

These three general approaches can also be combined, e.g. an article may delineate brain sites for one part of the brain as well as adopt brain sites from another map for a different part of the brain. Also, an article that only adopts brain sites can clearly state the source map for some brain sites and be ambiguous for other brain sites.

There is a fourth approach which, fortunately, is not often found in the literature and does not use brain sites according to our definition above. Instead, either broad topographical labels are used to describe data or experimental results are displayed by figures which lack names and borders of any brain sites and for which none of the map(s) defined or adopted by the article is valid. In CoCoMac, such purely topographical descriptions are not included into the brain map for a given article. Instead, for representing data described in a purely topographical way, a brain site from a suitable map is chosen whose location matches the topographical description of the label. As an example, for the rather vague description '...in the prefrontal cortex anterograde transport was found in the ventral bank of the principal sulcus and in the cortex lying immediately lateral to it...' (Matelli *et al.* 1986, p. 286) one would choose a suitably defined area that includes the principal sulcus (e.g. Walker's (1940) area 46). In order to highlight that this representation is based on a subjective (and thus error-prone) interpretation of a topographical description, (i) a comment is added to this entry explaining the reason for choosing the particular brain site, and (ii) a particular code

Table 2. Twelve different types of *BrainSite* representation objects that are currently used within CoCoMac for the representation of data referring to various ‘modules’ of brain architecture.

(Note that this scheme can be flexibly extended, e.g. to subcompartments of single neurons.)

object name	used for	definition of EC	representation of laminar data
<i>BS_IsoCtx_2D</i>	isocortical areas	projection from three- to two-dimensional	represented by intrinsic property
<i>BS_IsoCtx_3D</i>	isocortical areas	three-dimensional	represented by intrinsic property
<i>BS_AlloCtx_2D</i>	allocortical areas	projection from three- to two-dimensional	represented by <i>BS_Lamina_Ctx_2D/3D</i>
<i>BS_AlloCtx_3D</i>	allocortical areas	three-dimensional	represented by <i>BS_Lamina_Ctx_2D/3D</i>
<i>BS_Lamina_Ctx_2D</i>	cortical laminae	projection from three- to two-dimensional	represents a single lamina
<i>BS_Lamina_Ctx_3D</i>	cortical laminae	three-dimensional	represents a single lamina
<i>BS_Column_Ctx_3D</i>	cortical columns	three-dimensional	—
<i>BS_Band_Ctx_3D</i>	intracortical volumes	three-dimensional	—
<i>BS_Nucl_SubCtx_3D</i>	subcortical nuclei	three-dimensional	represented by <i>BS_Lamina_SubCtx_3D</i>
<i>BS_Lamina_SubCtx_3D</i>	subcortical laminae	three-dimensional	represents a single lamina
<i>BS_Neuron_Soma_3D</i>	neuron somata	three-dimensional	—
<i>BS_Neuron_Ensemble_3D</i>	ensembles of neurons	three-dimensional	—

(*PDC_Site* = P, see § 2b(iv)) is assigned to this entry, also pointing to the fact that the identification of the labelled brain site is somewhat speculative. At a later stage, this code allows us to treat the statement with low priority when it is necessary to compare and integrate contradictory statements from different articles (see § 2c(ii)).

Before data of a new article ‘X’ are entered into CoCoMac, one first needs to analyse what type of brain sites are used by X to describe its data. First, delineated brain sites are not only used by X itself to assign brain locations to its findings, but are likely to be specifically adopted by subsequent articles for a description of their results. The representation of delineated brain sites in CoCoMac thus is a fundamental precondition for the representation of all experimental data. Second, if X unspecifically adopts a brain site A, we are not entirely sure about the origin of A, either because no reference is given at all or because ambiguous references to several other articles coexist. We thus strictly follow our axioms (i) and (ii) in the list in § 2a by (i) creating a representation of A as a brain site of X according to the article’s nomenclature, and (ii) by explicitly documenting our assumptions about inter-map relationships of this brain site that can subsequently be used for mapping by ORT. Transparency of any assumptions is achieved: (i) by writing a commentary in which the assumption is explained in detail and (ii) by assigning an appropriate ‘precision of description’ code (PDC) that reflects the degree of subjective interpretation (see § 2b(iv)). In this way, data representation and data analysis are not being intermingled, we maintain a 1:1-representation, and all assumptions that are used for subsequent analyses or transformations are completely documented and transparent. As a third and final case, if X specifically adopts a brain site A from another article ‘Y’, we do not need to create a representation of A in X. Instead, a representation of A in Y is necessary before the data of X that refer to A can be represented in CoCoMac.

BrainSites form the most fundamental objects for data representation in CoCoMac. In other words, *BrainSites* are the basic ‘containers’ for all information that is stored by CoCoMac. For example, injection sites, labelled sites as well as densities,

laminar patterns and quantitative data of transported label are all described by reference to particular *BrainSites*, while the representation of *InterMapRelations* requires a pair of *BrainSites*. *BrainSites*, however, are not one totally uniform class of representation objects. CoCoMac currently makes use of 12 different types of *BrainSite* representation objects for various components of the brain that vary with respect to (i) the spatial dimension by which the extent of injected tracer or transported label is described and (ii) the representation of data on laminar patterns (table 2). These differences between the various types of *BrainSites* used within CoCoMac, however, are not important for an understanding of the general concept of data representation in CoCoMac, and will thus not be explored further here. Some more detailed information can be found in Appendix A. The main line to keep in mind for the next paragraphs simply is that *BrainSites* are the most basic objects of data representation by which all experimental data are stored in CoCoMac.

(iii) *Experiment-related data*

As can be inferred from tables 1 and 3, the class of experiment-related data in CoCoMac can be divided into three subgroups: tracing methods, injections and labelled sites. How are these subgroups connected? First, in each article, more than one tracing method may be presented by the authors. For example, the authors may use different tracer substances or different injection methods. For each tracing method, one or several experimental injections are described. These injections lead to more or less widespread label throughout the brain that results from active antero- or retrograde intraneuronal transport after the tracer substance has been taken up by somata or terminals, respectively. For each injection, the article thus describes a set of distinct brain sites in which label was found or was explicitly stated to be absent (note that the finding of a non-existing projection is equally important to the finding of an existing one). For each labelled site, the article may describe further properties of the results, e.g. its laminar pattern or quantitative data. This chain of hierarchical relationships is reflected by the relational structure of the respective tables in CoCoMac: the tables *Literature_Main*, *Methods*, *Injections*, *Labelled Sites_Descriptions*, and *LabelledSites_Data* are linked in descending

Table 3. Tables containing experiment-related data in CoCoMac. 'PDC' refers to a coding system for the precision of data description (see section 2b(iv)).

table	information on
<i>Methods</i>	tracer substances injection method (pressure, iontophoresis or diffusion) thickness of histological sections references to text and figures
<i>Methods_Animals</i>	collator's comments on tracing method number of animals used for the study macaque species age range weight range sex survival time
<i>Methods_Radioactive Tracers</i>	mixture of radioactive amino acids activity and specific activity of radioactive tracers
<i>Injections</i>	injected brain site <i>PDC_Site</i> of injection hemispheric location of injection (left or right) extent of injected substance <i>PDC_EC</i> of injection volume of injected substance concentration of injected substance affection of multiple brain sites by the injection methodological problems of the injection references to text and figures citations from the original text collator's comments on injection
<i>Injections_AffectedSites</i>	brain site additionally affected by an injection <i>PDC_Site</i> of additionally affected brain site extent of label in additionally affected brain site <i>PDC_EC</i> of additionally affected brain site
<i>Injections_Laminae</i>	laminae affected by injection <i>PDC_Laminae</i> of injection
<i>LabelledSites_Descriptions</i>	hemispheric location of label (ipsi- or contralateral) labelled structures (somata or terminals) references to text and figures citations from the original text collator's comments on description of label
<i>LabelledSites_Data</i>	labelled brain site <i>PDC_Site</i> of labelled brain site extent of label <i>PDC_EC</i> of label density of label <i>PDC_Density</i> of label
<i>LabelledSites_Data_Laminae</i>	collator's comments on labelled brain site laminar pattern of label within a labelled site <i>PDC_Laminae</i> of label
<i>LabelledSites_Data_QD</i>	labelled neurons per brain site (total number) labelled neurons within one brain site (%) labelled neurons within one lamina (%)

order by 1:n-relationships (see figure 1). The contents of the tables with experiment-related data are summarized by table 3. We now describe the representations of these data in more detail, starting with the *Methods* table.

Methods

The methods used for tract tracing vary in several respects. The most obvious variation concerns the choice of the tracer substance. A variety of different substances are available, classical ones like tritiated amino acids (TAA) and horseradish

peroxidase or the more recently introduced fluorescent dyes like fast blue and diamidino yellow. Each tracer substance has characteristic properties concerning the ratio and place of neuronal uptake (cell bodies or terminals), direction selectivity of intra-neuronal transport (antero- and/or retrograde), transport kinetics, sensitivity and selectivity of histological methods for detecting transported label (for a recent review see Köbbert *et al.* 2000). A second main factor of methodological variance is the injection method by which a tracer substance is applied to brain tissue. Often, this is accomplished by pressure injections using a

Figure 1. This screenshot shows 23 (out of a total of 52) tables in CoCoMac as well as their relational structure (compare tables 1 and 3). Bold field names of tables represent (components of) primary keys. Lines between tables represent relationships between primary and foreign keys. '1' and '∞' denote the 1- and *n*-sides, respectively, of 1:1- and 1:*n*-relationships. The upper row of tables contain experiment-related data, whereas the lower row of tables contain mapping data. Due to limitations of space, some tables are only shown by their titles without any field names.

Hamilton syringe. Alternatively, iontophoresis or diffusion of tracer substance from gel pads can be used.

Data on tracer substances and injection methods are usually available for each report and are thus contained by the *Methods* table (see table 3). Unfortunately, much less information is normally available on details of the histochemical procedures that are used to evaluate the experiments and which can also add substantial variability to the results of tracer experiments. At the moment, CoCoMac therefore does not contain specific tables for representing information on histochemical post-processing. Such tables can be easily created and linked to the *Methods* table, however, as soon as such information becomes available (for example, from the authors themselves). For the future, we also consider to store data on the physical locations of the needle tracks, i.e. whether the injected area was directly targeted (e.g. most cortical areas) or whether the needle had to penetrate other brain structures first (e.g. subcortical brain sites).

There are two further domains of methodological interest that are contained by two subordinate tables linked to the *Methods* table: information on the animals used (*Methods_Animals*) and, if used in the study, information on radioactive tracers (*Methods_RadioactiveTracers*). Information on animals is often surprisingly limited in the literature. Whereas data on the macaque species (e.g. *Macaca mulatta*, *M. fascicularis*, etc.), the number of animals used, and the survival time after injection are usually given, there is a frequent lack of information about age, sex and weight of the animals. As for radioactive tracers,

CoCoMac also contains data about activity, specific activity and the ratio of mixed different TAA (see table 3).

By representing each of these methodical details, CoCoMac allows the user to restrict queries or analyses to specific subsets of all data in CoCoMac that meet certain methodical requirements. For example, users can choose to extract only those reports from CoCoMac that were based on the use of fluorescent dyes in male animals of the *M. mulatta* species. Others may simply want to exclude results obtained by particular tracers, particular injection methods, and so on.

Injections

There are several basic features that every article should address when describing a tracer injection. What brain site was affected by the injection? How far did the injection spread within the targeted brain site? Were any neighbouring brain sites additionally affected by the injection? In the case that a cortical area is injected, which laminae were affected by the injection? Was the injected brain site located in the right or left hemisphere? What were the volume and concentration of the injected substance? Were there any methodological problems that might potentially confound a straightforward analysis of the results, e.g. spread of injected tracer substance to the white matter?

CoCoMac represents all data concerning these questions (see table 3). We shall discuss some of these issues in more detail now. One central question is whether the injected brain site is affected completely or only partially by the tracer substance. This

distinction is important because it has been demonstrated repeatedly that different parts of architectonically defined brain sites may have distinct patterns of connectivity with other parts of the brain. For example, Bates & Goldman-Rakic (1993) found that injections of retrograde tracers into various prefrontal areas consistently only labelled the most anterior part of supplementary motor cortex (their cytoarchitectonically defined area SMA). Using a more fine-grained parcellation for this region, Luppino *et al.* (1993) demonstrated that injections within various parts of caudal supplementary motor cortex (their area F3 as defined by cytoarchitectonics and cytochrome oxidase histochemistry) also resulted in partially different patterns of connectivity (compare figs 3 and 5 of their article). Generally, experiments with only partial affection of the injected brain site can potentially deliver falsely negative results, i.e. projections whose existence is falsely denied. This also explains why different partial injections into the same structure can yield seemingly contradictory results (see § 2c(ii)). CoCoMac uses a simple coding system for information on the extent of injected substance (as well as for resulting label—see § 2b(iii)) within a given brain site: the extension codes (ECs). These codes comprise five cases:

- (i) EC (A) = N: injected tracer substance (transported label) is found in no part of brain site A.
- (ii) EC (A) = P: injected tracer substance (transported label) is only partially found within brain site A.
- (iii) EC (A) = C: injected tracer substance (transported label) is found throughout the complete extent of brain site A.
- (iv) EC (A) = X: injected tracer substance (transported label) exists within brain site A, i.e. due to lack of precise information it is unknown whether it can be found for subparts of A only, or even for its complete extent.
- (v) EC (A) = U: it is unknown whether and to what extent injected tracer substance (transported label) is found within brain site A.

The use of ECs within ORT has been described in detail before (Stephan & Kötter 1999; Stephan *et al.* 2000a).

As important as the extent of tracer substance within the target brain site is the question whether tracer substance spreads beyond this site, i.e. whether more than one brain site is affected. Results from such cases are difficult to interpret since it cannot be easily deduced for any given labelled brain site whether one or several of the injected brain sites contributed to the observed label (see § 2c(ii)). CoCoMac thus lists all additionally affected brain sites (plus their ECs) in the table *Injections_AffectedSites* (see table 3). A second subtable, *Injections_Laminae*, contains information on the laminae that were affected by the individual injections. Most projections have a very distinct laminar pattern with respect to both their source and target brain site, and these patterns have been interpreted to reflect functional properties of the projections (Felleman & Van Essen 1991; Rockland & Pandya 1979). The coding that we use for describing laminar patterns is explained in detail in Appendix B.

Remember that our axioms (i) and (ii) above (§ 2a) demand a strict separation of data representation and data analysis as well as a 1:1-representation of the original data. CoCoMac therefore represents all available data on injections regardless of whether their subsequent interpretability is questioned by methodological problems like multiple affected sites, spread of tracer substance to white matter, etc. It is carefully documented, however, if and which methodological problems are present. How these data are eventually analysed depends on the

individual notions of the user. Just as different anatomists gain varying interpretations of the same data, the user can define several parameters that determine how the individual reports are weighted and analysed (see § 2c(v)).

Labelled sites

Data on labelled sites are represented by four hierarchically linked tables: *LabelledSites_Descriptions*, *LabelledSites_Data*, *LabelledSites_Data_Laminae*, *LabelledSites_Data_QD* (see figure 1 and tables 1 and 3). There are two simple reasons that explain why representing all data on labelled sites within one table would not be a good solution. First, there are some pieces of information that equally relate to many labelled sites resulting from the same injection, e.g. information about the hemispheric location of label (ipsi- or contralateral to the injection), information about which subcellular structures were actually investigated for label (somata or terminals), and references to text and figures. If all data on labelled sites were represented in one table, these descriptive items would be represented redundantly for many labelled sites. On the other hand, there are some data that are not provided by all authors or even by the same author for all experimental findings, e.g. information about the laminar pattern of label or quantitative data on labelled neurons. Representing all data on labelled sites in one table, a large proportion of entries would contain blank fields for these data. Formally, both cases contravene the normalization rules for relational databases that ensure optimal data integrity and performance (see § 2b(v)). By distributing the data over several hierarchically linked tables, we have chosen a solution that conforms to these normalization rules. The general descriptive information that is equally valid for all labelled sites belonging to the same injection is stored in one table (*LabelledSites_Descriptions*) as a basis for the site-specific data in a second table (*LabelledSites_Data*). Data on laminar patterns and quantitative data, respectively, are also represented in independent tables and linked to the associated entries of *LabelledSites_Data*.

CoCoMac contains both coded and non-coded information on labelled brain sites (table 3). The latter class includes citations of the original descriptions and precise references to page and figure numbers. These entries are part of the look-up component of CoCoMac and help the user to explore quickly data in their original source from the literature. Coded data, on the contrary, are the basis for algorithmic analyses of the data. With regard to data on labelled sites, CoCoMac contains four main types of coded data: extent, density and strength, laminar pattern, and quantitative data of the labelled neurons.

The extent of label within a given brain site is represented with the help of the ECs that are also used for describing the extent of tracer substance within an injected area (see § 2b(iii)).

'Density' or 'strength of label' are relatively vague concepts that are nevertheless used by a large number of tracing articles in an attempt to grade the anatomical strengths of different projections. The rationale for this effort is the assumption that there is a relationship between the degrees of anatomical and functional coupling between brain sites. Although there is no universally agreed upon definition, and although the terms are often used somewhat interchangeably, there seems to be a subtle difference between 'strength' and 'density'. 'Strength' may be defined by the total number of labelled neurons in a particular brain site (e.g. Luppino *et al.* 1993), whereas 'density' refers to the spatial concentration of labelled neurons (i.e. number of neurons per area unit—for an example see Carmichael & Price (1995b, p. 644)). Most articles are not specific about these

concepts and treat 'strong connections' and 'dense connections' equivalently. In any case, the grading is rather coarse, and even if quantitative and objective methods are used (e.g. automatic cell counts) any threshold for defining these classes is arbitrary. Within CoCoMac, we chose a pragmatic approach for representing these data by including both reports on strength and density without making a strict representational distinction between them. Analogous to the convention in the literature, we distinguish between three different and rather coarse grades that are coded numerically by values between 1 (weak) and 3 (strong). Textual information about density or strength is interpreted using the following keywords: (i) 'weak', 'sparse' or 'light' label $\rightarrow$ density or strength = 1, (ii) 'moderate' or 'medium' label $\rightarrow$ density or strength = 2, (iii) 'strong' or 'dense' or 'heavy' label $\rightarrow$ density or strength = 3. Information provided by figures is particularly useful if the label is represented by symbols that clearly assign label within a given area to one of the three classes above (for an example see Carmichael & Price (1996, figs 7–15)). Figures that indirectly display the density of label by plotting individual labelled neurons require further interpretation by the data collator. If no other information on the strength or density of label is available, these figures are evaluated, but the high degree of subjectivity is indicated by tagging the resulting entry with an appropriate PDC (see § 2b(iv)).

As already mentioned above, information on specific laminar patterns of projections are highly interesting data. Several theories of hierarchical organization, particularly within the visual system, have been based upon such data (Felleman & Van Essen 1991; Hilgetag *et al.* 1996; Ungerleider & Mishkin 1982). Whereas laminar patterns were often classified as groups in the past (e.g. as 'ascending', 'lateral', or 'descending' projection patterns—see Felleman & Van Essen (1991)), CoCoMac applies a different principle: for each lamina, one alphanumeric placeholder indicates presence or absence of label as well as the relative strength of label compared with the other laminae of the same area. Thus, laminar patterns in isocortical areas (which, by definition, have six main layers) are usually represented by a six-character code. This does not apply, however, to brain sites whose laminar structure deviates from the isocortical scheme. For example, laminae of primary visual cortex as well as of allocortical and subcortical brain structures are represented as individual brain sites, using the appropriate *BrainSite* types shown by table 2. More details of the principles of representing laminar information in CoCoMac can be found in Appendix B.

There are three main lines of quantitative data that are meaningful in the context of tracing studies. First, some experimentalists count the absolute number of labelled neurons within each brain site that they investigate. Others extend this approach by determining the percentage of labelled neurons for each lamina in relation to all other laminae of these brain sites. Alternatively, one can state the percentage of labelled neurons within each investigated brain site in relation to all neurons that were labelled by the injection throughout the entire brain. For data from these groups, CoCoMac contains three subtables of *LabelledSites_Data* which, in their entirety, are addressed as *LabelledSites_Data_QD* in this article (see tables 1 and 3).

(iv) PDCs: coding the precision of descriptions

For analyses of large datasets, it is necessary to compare and integrate the results from many individual articles. Unfortunately, integrating data from different articles in the literature is often substantially difficult because of three main problems:

(i) use of incompatible brain maps by which authors describe their findings (the parcellation problem), (ii) use of different tracing methods (e.g. with respect to tracer substances and histological protocols) with potentially different degrees of sensitivity and specificity, and (iii) different levels of precision of data description.

The first problem—as has been described above (§ 2b(ii))—is equivalent to comparisons between data points in different microstructural–functional feature spaces. CoCoMac tackles this problem by means of ORT (see § 2c(ii) and § 2c(iii)).

The second problem is harder: although the general methodological principle of all tracing studies is the same, individual tracers differ from each other in selectivity of transport direction, kinetics of intraneuronal uptake and transport, as well as sensitivity and specificity of labelling neuron somata or terminals (Burns 2000; Köbbert *et al.* 2000). Furthermore, even within the same article, the interpretability of individual injections may be rendered difficult by methodological problems, e.g. affection of multiple brain sites by the same injection, spread of injected tracer to the white matter or experimentally induced lesions. Eventually, laboratory-specific procedures, e.g. differences in histological post-mortem processing, can add to the overall variation of connectivity data. All available details about tracing methods and injections of each individual article are carefully represented in CoCoMac (see § 2b(iii)). Subsequently, on the basis of this information, the user has multiple options to impose personal preferences upon analyses or queries within CoCoMac (see table 4 and § 2c(v)). For example, the user can set several parameters that determine how injections with multiple affected brain sites or methodical problems are evaluated. Alternatively, one can restrict queries or analyses to specific subsets of data in CoCoMac, e.g. by excluding all projections that were experimentally determined by particular tracer substances. Altogether, the user can thus look at different aspects of the same data by differentially weighting various methodical parameters.

The third problem, i.e. different levels of precision of data description, also is very commonly met when comparing data between and even within articles. Different authors follow different strategies for presenting their experimental results. To juxtapose two extremes of the spectrum, some authors precisely relate their findings of existent or absent label to carefully defined brain sites and provide additional figures with clearly delineated borders, clear names of brain sites, and clearly plotted label. Others may just refer to vaguely defined topographical regions and provide sketchy diagrams with approximate locations of label and without names and borders of brain sites. Generally, one can conclude that the less precise a description of data in the literature, the higher will be the variation between different 'observers' of these data with respect to constructing representations of these data within a database. In other words, the objectivity (i.e. observer-independence) of any method for 'observing' and interpreting data from the literature strongly depends on the precision of data description.

These problems become very important as soon as contradictory statements from different articles are compared: which of the conflicting data should be assigned higher priority? Ideally, one would like to determine data validity by some kind of 'truth index' for any given statement from the literature that is represented in CoCoMac, i.e. the probability that the described finding correctly describes the 'true' connectivity of the macaque brain. Decisions about conflicting reports could then be made on the basis of such indices. Unfortunately, there exist no clear

Table 4. Parameters of analysis that can be defined by the user, and their influence on selecting, integrating and transforming primary data within CoCoMac.

parameter number	name	influence on data analysis
1–5	PDC threshold (for each of the five PDC classes)	selection of primary projections
6–10	PDC hierarchy (for each of the five PDC classes)	selection of primary projections integration of inter-map relations construction of transformation graph integration of primary and transformed projections
11	target map	transformation of integrated data
12–15	methodological problems of injections that:	selection of primary projections and integration of primary projections
	affect several brain sites simultaneously	selection of primary projections and integration of primary projections
	spread to white matter	selection of primary projections and integration of primary projections
	are performed in lesioned brains	selection of primary projections and integration of primary projections
	apply same tracer to both hemispheres	selection of primary projections and integration of primary projections
16	mode of integration of projection strength	integration of primary and transformed projections
17	ipsi- versus contralateral projections	selection of primary projections
18	right versus left versus both hemispheres	selection of primary projections
19	tracer substances	selection of primary projections
20	macaque species	selection of primary projections
21	gender of animals	selection of primary projections

criteria by which the different sources of variation described above could be assessed objectively, and with which their relative importance should be weighted. Within CoCoMac we refrain from defining such a ‘truth’ index. Instead, we reformulate the above question for the ‘truth’ of data in the literature. Applying criteria developed by theoretical psychology, ‘truth’ of experimental data crucially depends on how these data are measured and represented by an observer. If the chosen method of observing and representing data fails to satisfy the fundamental measurement criteria of objectivity, reliability, and validity (see § 2a), it becomes very difficult to distinguish between ‘correct’ and ‘incorrect’ data since errors or variation due to the experimental design can no longer be distinguished from errors or variance due to the method of observation and representation. Furthermore, these three criteria possess a hierarchical inter-relationship: high objectivity is a necessary (although not sufficient) condition for high reliability, which again is a necessary (although not sufficient) condition for high validity (Bortz & Döring 1995; Guilford 1954; Lienert & Raatz 1994). Therefore, it appears useful to assess primarily the objectivity of the data in the literature. As already mentioned above (§ 2a), this is feasible by estimating for any given datum the likelihood that individual data collators would have interpreted it differently and thus created non-equivalent representations of it in databases. In all subsequent data comparisons and analyses, whenever one needs to choose between contradictory statements from different articles, one would prefer that statement which was described most precisely and which is thus least likely to be confounded by errors of interpretation.

As a scale for the descriptive precision of data, CoCoMac introduces a new coding scheme: PDCs. The main principle of PDCs is to define a set of clear criteria concerning data representation within each data modality. For example, the identification of a cortical area that is labelled (or not labelled) by transported tracer substance can be accomplished by textual

descriptions, tables, exact figures, diagrams, photographs, or by a combination thereof. The text may explicitly identify an area by its full name or acronym, or refer to it vaguely by topographical descriptions. Figures may or may not show clear areal borders, specific areal names, and the exact extent of tracer substance. Text and figures may confirm, complement, or contradict each other, and so on. Regarding each combination of such criteria as a specific case, one obtains a set of disjoint classes with different degrees of ambiguity (see figure 2). It is possible to define these criteria very carefully, so that the overall coding is well operationalized and thus standardized between different observers.

It should be noted that even within a given article, the precision of data description may vary considerably. For example, a particular article may very precisely identify the brain sites that were labelled after a particular injection, while providing rather ambiguous statements on the density or the laminar patterns of the label. Alternatively, the precision of data description may vary from finding to finding within the same class of items. Therefore, PD coding is not only applied once to each article contained by CoCoMac, but to every individual datum of the following modalities: (i) identification of injected and (un)labelled brain sites (*PDC_Site*), (ii) extent of tracer substance or transported label within an identified area (*PDC_EC*), (iii) density or strength of label within an identified area (*PDC_Density*), (iv) laminar patterns of label within an identified area (*PDC_Laminae*), and (v) inter-map relationships of two brain sites (*PDC_RC*). On the whole, PDCs can be conceptualized as a systematic observation method covering five categories.

How exactly are these PDCs defined and determined? As limitations of space prevent us from listing the rules for all five data modalities mentioned above, a representative example is shown by figure 2. It shows diagrammatically how the precision of a brain site’s identification (i.e. *PDC_Site*) can be

Figure 2. Decision tree for determination of PDCs with respect to the identification of injected, labelled, or unlabelled brain sites (*PDC_Site*). Starting at the top, the collator is faced with a series of clear binary decisions until a discrete case (i.e. a leaf of the tree, represented by an encircled capital letter) is reached. Note that although the individual cases are labelled by consecutive uppercase letters, the alphabetical order is not meant to imply any predefined hierarchy among the PDCs. Instead, the user can define variable hierarchies, depending on personal preferences.

determined according to clear rules. Somewhat different, but analogous rules exist for the other four PDC classes (see <http://www.cocomac.org> for more information). It should be noted that although we label the individual cases by consecutive uppercase letters, the alphabetical order is not meant to imply any predefined hierarchy among the PDCs. Instead, the user can define variable PDC hierarchies, depending on personal preferences. For example, before starting the analysis routines of CoCoMac, one might assign a higher value to data based on textual descriptions than on data merely displayed by figures and thus rank the PDCs accordingly. If contradictions between text and figures should be met during data analysis, the text-based data would win over the figure-based data. Other users may have the opposite preference and thus choose a different PDC hierarchy. Therefore, whenever we speak of 'better' or 'higher' PDC, we do not define superiority or inferiority in absolute terms, but always relative to a given user-chosen hierarchy.

Generally, PDCs are thus a means of indirectly estimating the validity of data from the literature at the early stage of objectivity (defined as observer-independent representation). As will be described below (§ 2c(ii), 2c(iv)), PDCs are an important element for all algorithms used by CoCoMac to analyse and integrate its data according to the user's preferences.

(v) Relational structure of CoCoMac

As far as data representation is concerned, CoCoMac is a relational database system. The relational model has been one of the most successful paradigms in database technology and is

based on a large and mathematically solid body of theory (e.g. see the classic paper by Codd (1970), or for a modern textbook Ullman & Widom (1997)). The core idea of the relational model is that data are represented as entries in distributed, independent tables that are linked by hierarchical relationships. Each table contains one designated field (or a designated set of fields) for which each entry has a unique value (or a unique combination of values)—the primary key of the table. Single fields or a combination of fields that are linked to a primary key of another table and thus unambiguously cross-reference tables, are called foreign keys. Several examples of such hierarchical relationships between tables are shown in figure 1.

For any given set of data, there are various ways of representing them within a relational database. As the different options vary enormously with regard to memory demands, performance of data processing and data integrity, research on relational databases has focused very early on optimally efficient approaches to data representation. The central result of these efforts are a series of 'normalization rules', first formulated by Codd (1970). The relational structure of CoCoMac reflects these normalization rules and thus ensures data integrity and good performance.

The tables of CoCoMac can be grouped into three main classes: bibliographical, mapping, and experiment-related data (table 1). The tables of bibliographical data form the basis of the relational hierarchy in CoCoMac from which two distinct streams of tables emerge (figure 1). One stream (lower half of figure 1) consists of those tables that contain mapping data (e.g. *BrainMaps*, *InterMapRelations*) whereas the other one contains all

tables with experiment-related data (upper part of figure 1), e.g. information on tracing methods, experimental injections, and labelled brain sites. The two streams are interconnected by relationships between the table *BrainMaps_BrainSites* and all experiment-related tables whose data require a reference to brain sites.

(vi) *Standardized principles of data collation and representation*

To standardize data collation and data representation in CoCoMac between different observers we have defined a set of clear rules which define the order and contents of the various aspects of data collation and data representation. Some of these rules have been mentioned at various sections of this paper, e.g. they determine whether a given brain site needs to be explicitly represented within CoCoMac, they define the coding schemes for extent, strength and laminar pattern of label, specify the criteria according to which a PDC is chosen for a given datum, etc. These rules are compiled in a manual comprising more than 50 pages and are thus too numerous and detailed to be all explicitly listed within this paper's but are available on the Web via <http://www.cocomac.org>.

Since data representations in CoCoMac are created by several people's we have been able to empirically verify the trainability of our scheme. Different observers who are trained to represent published data in CoCoMac according to our rules indeed manage to achieve equivalent results. As part of this standardization, we have begun a proofreading process in which representations of data from a given article that have been entered by one person are checked for compatibility with our data collation rules at a later time, preferably by another person, and corrected if necessary. We hope that linking CoCoMac to the internet will also encourage the original authors of the included articles to take a more active part in validating the data representations of CoCoMac.

(c) *Data analysis*

(i) *General concepts of data analysis and integration in CoCoMac*

We distinguish three different classes of data in CoCoMac: primary, integrated, and transformed data (see left column in figure 3). Primary data are those entries of CoCoMac that were created by a data collator to represent original experimental findings described by the literature. They constantly exist in the database without being altered by any analysis. For example, Carmichael & Price (1995a, pp. 617–619) describe an injection into the 'intermediate agranular insular area' (area Iai) as defined by their own parcellation (Carmichael & Price 1994), with the injection being placed into the right hemisphere, affecting only laminae I–III and extending only to subparts of area Iai. The primary data in CoCoMac corresponding to this statement are created by setting the *BrainSite* field of the *Injections* table to the value 'CP94-Iai', the *EC* field to 'P', the *Hemisphere* field to 'R', and the *Laminae* field of the *Injections_Laminae* table to 'XXX000'.

In contrast, integrated data result from analysis and comparison of redundant or contradictory primary data. Integration is performed according to a series of parameters, which can be defined by the user (see §2c(v)). As a simple example for integration of primary data, we consider two statements from different articles that contradict each other with regard to the existence of a particular projection $A \rightarrow B$ (statement 1: 'brain site A projects to brain site B'; statement 2: 'brain site A does not project to brain site B'). How can we integrate such conflicting information? A closer look at the primary data might reveal

that statement 2 was based on an injection that affected only a subpart of A (i.e. $EC_{injection2}(A) = P$), whereas statement 1 was derived from an injection that covered the complete extent of A (i.e. $EC_{injection1}(A) = C$). Assuming that both statements precisely identified brain sites A and B (i.e. they were assigned similarly high *PDC_Site*), one could then conclude that the apparent contradiction between the two statements does not really exist, but that statement 2 is simply restricted to a subpart of A which does not send fibre connections to B, whereas the remaining part of A does project to B. If, however, statement 2 precisely identifies the involved brain sites A and B and unambiguously states that B is not labelled after partial injection of A, whereas statement 1 is too vague to decide whether its information really refers to the same projection $A \rightarrow B$, one would hesitate to come to the same conclusion. Rather, one might prefer to conclude that the projection $A \rightarrow B$ is indeed absent since statement 2 delivers the only securely interpretable information about it. More details on principles of data integration are described in §2c(ii) and §2c(iv).

Finally, transformed data result from mapping integrated data by ORT (see §2c(ii) and §2c(iii)) to a user-chosen target map. Like integrated data, the computation of transformed data depends on the underlying primary data as well as on several user-chosen criteria (e.g. PDC hierarchies, target map). In other words, the results of integrating or transforming data are different as soon as either the primary data or the user-specified criteria change. For this reason, integrated and transformed data are not kept as constant entries by CoCoMac but are computed anew when data or criteria have changed.

For each class of data in CoCoMac, there is one associated class of analysis type: simple queries, meta-analyses, and network analyses (see right column of labels and dashed arrows in figure 3). The first class contains simple queries that extract user-defined subsets of primary data from CoCoMac. Since CoCoMac is a relational database, these queries are easily implemented by means of SQL (structured query language) commands. For example, users may wish to view all primary experimental data that were obtained by one specific injection or, vice versa, list all injections that delivered information on one specific projection. Other users may want to extract all primary data that were delivered by particular tracing methods, e.g. to juxtapose the results obtained by anterograde tracers like TAA to those gained by retrograde fluorescent dyes. Generally, within the frame of SQL syntax, arbitrary combinations of fields from any of the tables of CoCoMac can be easily linked by Boolean expressions (e.g. logical AND, OR, NOT) to allow for powerful and individualized data extraction from CoCoMac.

For many purposes, however, simple SQL queries of the above type are of limited value. If primary data are described within the same brain map, they simply juxtapose redundant and contradictory findings without integrating them. On the other hand, if primary data are described by different parcellation schemes, the equivalence of findings described across incongruent maps is not easily recognized by such simple queries. Therefore, meta-analyses that operate on primary data integrated within maps or on integrated data that have been transformed to a common target map by ORT are more powerful. This class of analysis makes it possible to address a very broad range of questions on brain connectivity. For example, one can directly contrast the connectivity patterns between individual brain sites or compare the ipsi- versus the contralateral projection pattern of any given brain site. Researchers interested in lateralization of brain functions can extract connectivity data for the left and the

Figure 3. Different levels of data (left uppercase labels) in CoCoMac and associated types of analysis (right lowercase labels). Note that meta-analyses and network analyses can be performed both on the basis of integrated and transformed data.

right hemisphere separately to compare their connectivity patterns. The influence of sex on brain connectivity can be investigated by comparing connectivity patterns between female and male animals. Any single or several of these questions in combination can be further related to individual authors, particular macaque species, specific tracing techniques, and so on.

The third class of analysis, network analysis, also requires at least integrated data, but is more powerful when applied to matrices of connectivity data that have already been transformed to a common brain map. Since the implementation of ORT, it has been possible to produce and then export such connectivity matrices from CoCoMac into any other software application for subsequent analysis (e.g. see Bozkurt *et al.* 2001; Northoff *et al.* 2000; Hautzel *et al.* 2000; Kötter *et al.* 2001a,b; Stephan *et al.* 2000b). We are currently beginning to implement several mathematical tools so that various analyses can be performed directly within CoCoMac. These tools include graph-theoretical algorithms for computation of shortest projection paths within arbitrary sets of brain sites or cluster indices (Watts & Strogatz 1998) as well as algorithms that compute some recently introduced indices for biological neural networks (Kötter & Stephan 2000). Also, we are planning to introduce interfaces to other software packages that use connectivity data, e.g. the knowledge base NeuroScholar (Burns, this issue) or the CANTOR system (O'Neill & Hilgetag, this issue).

The next sections focus on the transitions between primary, integrated and transformed data (see bold arrows in figure 3). Due to restrictions in length and scope of this paper, we do not explain mathematical details of data integration and transformation, but only give an outline of the basic principles. We further explain some of the parameters that the user can set in order to influence data integration and transformation. In general, we distinguish three main stages of the process from primary to transformed data (figure 4) that we shall explain in the following sections.

(ii) *Preprocessing of primary data*

Before CoCoMac begins to transform data to a user-chosen output map, it performs several steps of preprocessing on the

primary data (see figure 4). Briefly summarized, preprocessing fulfils three major requirements: first, to decide which of the primary data are actually methodologically acceptable according to the user-chosen criteria; second, to integrate contradictory and redundant primary data into a consistent set of data; third, ORT constructs an optimal transformation graph, which is subsequently used for transformation of the integrated data.

Integration of inter-map relationships

As a first step of preprocessing, CoCoMac searches the table *InterMapRelations* for redundant and contradictory statements that need to be integrated. For example, the literature contains two incompatible notions of the relationship between area 9 of Walker (1940) and area 9 of Brodmann (1905). Whereas Von Bonin & Bailey (1947, p. 66) rather vaguely imply that the two areas only partially overlap, Petrides & Pandya (1994) clearly come to the conclusion that W40-9 is a subarea of B05-9 (p. 27: 'In Brodmann's (1905) map of the monkey brain, ... no area 46 was identified ... By contrast, Walker (1940) designated a large portion of the region labelled as area 9 by Brodmann ... as area 46, limiting the designation area 9 only to the dorsal part of this region.') The decision among such incompatible notions (which are, however, rare and currently account for less than 1% of the relationships contained by CoCoMac) can be performed in two different ways. Either CoCoMac algorithmically compares the contradictory statements with respect to the precision by which they describe the inter-map relationship (i.e. on the basis of the associated *PDC_RC* and their user-defined hierarchy), or the user can manually decide for any conflict which inter-map relationship should be preferred.

ORT: constructing an optimized transformation graph

After contradictory and redundant inter-map relationships (and their PDCs) have been integrated, CoCoMac invokes the first component of ORT, i.e. the construction of the so-called 'transformation graph'. The computational principles of this step are derived from an 'all-pairs-shortest-paths' graph algorithm originally described by Floyd (1962). Our modified version of

Figure 4. Summary of the three main stages of data processing from primary to transformed data.

this algorithm has been described in mathematical detail before (Stephan *et al.* 2000a), thus we here only briefly summarize its basic principles. Initially, our algorithm represents all *BrainSites* and *InterMapRelations* as a graph, i.e. as a set of nodes connected by a set of edges. Two nodes are connected by an edge if there is an *InterMapRelation* (represented by an RC) between the respective *BrainSites*. A path within this graph then simply is a sequence of *InterMapRelations* (and can thus be represented as a sequence of RCs). Our algorithm operates on this graph with the aim of iteratively constructing all valid and optimal transformation paths between all *BrainSites*. In our context, 'valid' means that a transformation path allows unambiguous logical mapping of data along its sequence of *InterMapRelations* by means of the algebra operations of ORT (see § 2c(iii)). Each valid pathway is logically equivalent to one of the four non-disjoint RCs (I, S, L and O, see § 2b(ii) above). 'Optimal' means (i) that the path contains *InterMapRelations* that are as precise as possible (i.e. possess optimal *PDC_RC*), and (ii) that the overall logical mapping along the path is as simple as possible. Each possible combination of pathways within the graph is evaluated for these two criteria by means of a finite automaton. Overall, our algorithm thus not only determines the most precisely described and simplest transformation paths, but also deduces previously non-described inter-map relationships from the available statements in the literature. A quantitative example illustrates the power of this approach: the initial 3523 inter-map relationships from the literature that were contained by CoCoMac on 30 April 2000, were expanded by the algorithm to 91 055 inter-map relationships in the form of transformation paths.

The construction of the eventual transformation graph, however, requires yet one further step. Although each of the paths delivered by the above procedure is logically valid as such ('within-paths consistency'), there nevertheless may be logical contradictions between different paths that have one *BrainSite* in common ('between-paths inconsistency'). These mathematically definable contradictions are algorithmically identified and resolved on the basis of their PDCs and RCs (for details see Stephan *et al.* 2000a). The transformation paths that remain after this sieving procedure (65 964 in the example mentioned above) can subsequently be used by the algebra operations of ORT for mapping data in a logical, non-spatial manner between any pair of parcellation schemes (see § 2c(iii)).

Composition of primary projections

As the next step of preprocessing, we need to address which of all primary data should actually be accepted for subsequent integration and transformation. This decision is necessary because there are some frequent methodological problems that may limit the interpretability of experimentally observed label. Since different attitudes to these problems are found in the literature, CoCoMac does not rely on a fixed set of predefined rules but allows its users to interpret the primary data from different perspectives. Generally, for all methodological uncertainties, the user can determine precisely whether and how the affected data should be interpreted. The user's preferences of how data are selected and interpreted are represented by a set of parameters that must be defined before each analysis and are saved for documentation together with the eventual results (see § 2c(v)).

There are several reasons why a particular finding may not be considered acceptable for analysis. For example, the user may wish to analyse or transform only those data that were described by the literature with a certain minimum of precision. Within CoCoMac, one can formally define such a personal threshold by means of the five PDC hierarchies and specify that all data with PDCs below these thresholds are not taken into account for analysis and transformation. Beyond subjective preferences, the interpretability of certain data may be difficult due to methodological problems of the respective injection. In the case of injections that showed spread of tracer substance to white matter, for example, it can sometimes not be excluded that parts of the observed label may have been produced by the uptake of tracer substance by passing fibre tracts. Other methodological problems comprise experimentally induced lesions or simultaneous injections of the same tracer in both hemispheres under the (often untested) assumption that the injected brain site does not possess commissural projections.

The most frequent and important methodological problem, however, is that of an injection that affects several brain sites at the same time. We shall therefore describe this problem and the different approaches that CoCoMac offers to its solution in more detail. The problem with an injection that simultaneously affects several brain sites $A_1, A_2, \dots, A_n$ is that we cannot differentiate which of the injected areas contributed to observed label by means of its connectivity. In analogy to the varying attitudes

among anatomists towards the interpretation of such injections, CoCoMac offers the user the choice between a strict (STRICT), a liberal (LIB), and an intermediate (MED) interpretation of data from such injections. The STRICT approach does not assign the observed label to any of the original brain sites $A_1, \dots, A_n$ affected by the injection. Instead, by using information on the inter-map relationships between brain sites, it searches for a set of 'supra-sites' Σ from arbitrary brain maps so that each brain site P from Σ comprises all of the brain sites $A_1, \dots, A_n$ affected by the injection:

$$\Sigma = \{P \mid \text{RC}(P, A_i) = L, i \in \{1, \dots, n\}\}. \quad (1)$$

The projections described by the experiment can then adequately be restated on the basis of these supra-sites. Although this procedure diminishes resolution of the data, it is the prerequisite for their correct identification. As a real-life analogy, although one may have conflicting information whether a particular person is from London or Manchester, one can still deduce that this person is from England. As an example from neuroanatomy, Barbas & Pandya (1989, p. 366) describe an injection that simultaneously affected their areas V46 and D46. As both these areas are included by Walker's (1940) area 46, the latter represents one possible supra-site for this injection. As Barbas & Pandya found anterograde label in their area D10 after this injection, the primary projection can be adequately restated as W40-46BP89-D10 (note that this is not the representation of the authors' finding within CoCoMac but an inference made during the analysis only). Another potential supra-site would be Brodmann's (1909) area 9 that includes BP89-V46 and BP89-D46 as well. However, W40-46 also is a subarea of B09-9, therefore all information described by means of W40-46 for this injection can be easily mapped to B09-9. Therefore, we would not gain any further information by additionally representing the results of this injection on the basis of B09-9. Generally, CoCoMac thus computes an optimized set Σ_{opt} of smallest supra-sites:

$$\Sigma_{\text{opt}} = \{P \in \Sigma \mid \neg(\exists Q \in \Sigma : \text{RC}(P, Q) \in \{I, L\})\}. \quad (2)$$

If STRICT is not able to find any supra-site for a given injection, information on labelled brain sites provided by this injection on existing label is not evaluated at all. However, it is still possible to use information on label whose absence has been explicitly stated. For example, if a brain site B is clearly described to show no label after an injection of anterograde tracer simultaneously affected brain sites $A_1, \dots, A_n$, it is logically correct to deduce that none of these brain sites (at least not their subparts that were affected by the injection) sends a projection to B .

The same approach for absent label is used by LIB to evaluating injections with multiple affected brain sites. In contrast to the stricter approach described above, however, it does not make the effort of looking for supra-sites, but simply assigns the observed label to the brain site that was predominantly affected by the injection (an approach that is indeed pursued by some authors in the literature).

Finally, there is MED that combines the principles of the STRICT and the LIB approach by assigning label to the predominantly affected brain site only if it has not been possible to determine a supra-site for the simultaneously injected brain sites.

The user can choose between these three approaches and thus decides which of the primary data are accepted for subsequent

integration and transformation. Although only the strict approach effectively prevents misinterpretations of the data in terms of falsely positive projections, it may still be warranted to choose the MED or LIB approach if one wants to analyse data on brain regions for which only very few experimental data exist in the literature.

After it has been decided on the basis of the user's preferences which of all original data are accepted for integration and transformation, CoCoMac composes the primary projections by selecting and combining data from various tables (compare table 3). In a next step, these primary projections are integrated into a consistent data set.

Integration of primary projections

Due to the fact that many brain sites have been injected several times within and across studies, there is a considerable amount of both redundant and contradictory information in the literature. Thus, for each primary projection that has been described more than once in the literature, we need to integrate the various reports, i.e. to condense them into a single piece of information that reflects an overall interpretation of all available data. Without going into mathematical details, we describe three main aspects that guide this process of integration.

First, one would like to judge competing reports on the basis of their 'truth'. As a measure of 'truth' is not readily available, CoCoMac makes use of the PDCs as a measure of the precision of data description. The rationale for this is that the precision of data description is a main determinant of the objectivity (observer-independence) of data collation from the literature into CoCoMac, and high objectivity is a necessary condition for high reliability and validity (see §2b(iv)). Although the PDC of a given datum therefore does not allow us to deduce directly its validity, it nevertheless constrains the probability of the datum being valid. In the context of comparing two competing primary projections P_1 and P_2 (both of which describe a projection $A \rightarrow B$), the fundamental question is whether in both reports the identification of the two brain sites A and B was described sufficiently precisely that one can be certain that the information of P_1 and P_2 really relates to $A \rightarrow B$. More formally, if either $\text{PDC_Site}(A)$ or $\text{PDC_Site}(B)$ of P_1 is less precise than a user-chosen threshold, one would no longer assume with confidence that the information of P_1 really concerns $A \rightarrow B$ (the respective conclusion applies to P_2). Evaluating this question for both P_1 and P_2 , one obtains a first criterion to decide which primary projection should be preferred.

Second, it is important to note that, except for intraneuronal injections, a single injection cannot clarify whether neurons in the entire region affected by tracer substance or only in a subpart of it contributed to transporting label to a particular distant site. With respect to a single injection that delivers data on a given projection $A \rightarrow B$, the information on the injected site is of binary nature: either some (unknown) part of the injected site sends fibre projections to the labelled site or it does not. Specifically, injections of anterograde tracers in the source site A provide such binary information on the source site A , whereas more detailed information on the extent of label in target site B is available (the opposite constellation is present for injections of a retrograde tracer in the target site B). This means that two primary projections based on different tracing directions are only of limited direct comparability since their corresponding ECs bear somewhat different information. CoCoMac takes these considerations into account: when integrating a set of primary projections, it first subdivides these projections into two groups

based on antero- and retrograde tracers, respectively, and integrates these groups separately before then integrating the results from the two groups.

Third, we need to consider the actual extent of the injections within the injected site (i.e. $EC_{injection}$). There are numerous examples of contradictory findings in the literature where, after partial injection into a brain site A, one author reports label in another brain site B, while a different author whose injection covered a different subpart of A cannot confirm this finding (see figure 5). This inconsistency often seems to be due to the fact that common methods for experimentally delineating a given brain site (e.g. by cytoarchitectonical criteria) do not guarantee that all subparts of this brain site possess an identical pattern of association fibre connectivity. Various authors have reported obvious connectivity differences among subparts of architectonically defined brain sites. Consequently, it is not necessarily a contradiction if one injection into a given brain site A yields label in another brain site B, whereas another injection into a different subpart of A does not lead to any label in B. As long as one relates the description of the projection to the entirety of A and not to its subparts, one will therefore let the existing projection 'win' over the absent projection. This principle is also applied in CoCoMac—although constrained by additional consideration of the precision by which the brain sites were identified (i.e. the *PDC_Site* of all brain sites constituting the projections).

Such conflicting reports of differential connectivity patterns within architectonically defined brain sites challenge current concepts of what constitutes a homogenous brain site as an independent 'module' of the brain (Gahr 1997). This problem no longer exists, of course, for those injections that affect the injected structure completely (i.e. whose $EC_{injection} = C$). Such injections will reveal the combined connectivity patterns of all subparts (see figure 5d). Consequently, CoCoMac not only takes into account information on the descriptive precision and tracing direction, but also evaluates whether any of two compared injections covers the entire extent of the injected site.

(iii) *ORT: transformation of integrated projection data to user-chosen target map*

After the primary projections have been integrated, the next step is to transform them to the user-specified target map. This step is accomplished by ORT. The mathematical principles of ORT have been described in full detail before (Stephan & Kötter 1998, 1999; Stephan *et al.* 2000a), therefore we here only briefly summarize the main ideas. The basis of ORT is formed by an algebra of transformation, i.e. a set of operations on the two sets of ECs and RCs (see § 2b(ii) and § 2b(iii)). These operations represent simple rules that explain how non-spatial information on nominal scale (ECs) can be converted between different parcellation schemes depending on the logical relationships (RCs) between them.

As the literature does not provide statements on the inter-map relationships for all possible pairs of brain sites from different maps, it is necessary to deduce information on unknown inter-map relationships from known ones. Therefore, an essential component of ORT is to use a combination of graph-theoretical techniques, formal languages, and finite automata to construct a transformation graph that contains all possible transformation paths between the brain sites known by CoCoMac which are logically valid and have optimal mapping qualities (see § 2c(ii)). The algebra operations of ORT are then applied to these transformation paths to convert ECs between brain sites of different maps.

Figure 5. Some diagrammatic examples for partial and complete injections that show how different injections into the same brain site can lead to seemingly contradictory findings. (a) 'True' connectivity between two virtual brain sites. Only a subpart of the source site S sends association fibres to a subpart of the target site T. (b) If partial antero- and retrograde injections are placed within those subparts of the two brain sites that send and receive association fibres, respectively, the experimental results correctly point to an existing projection. (c) If, however, partial antero- and retrograde injections are placed into those subparts of the two brain sites that do not send and receive association fibres, respectively, the experiment falsely indicates that the projection does not exist. (d) Antero- and retrograde injections that affect the complete extent of the source site S and the target site T, respectively, deliver secure information in any case.

It should be mentioned though, that in spite of all optimization, mapping results obtained by ORT are not always unique since the initial information is non-spatial (i.e. does not rely on coordinates): for some combinations of ECs and RCs, the transformation gives ambiguous results. Within CoCoMac, ambiguous results at any stage of the mapping process are generally abandoned and not used any further, thus the final output from any transformation results from unambiguous operations only.

On the whole, ORT represents a fully formalized, algorithmic method that allows for effective, reproducible and transparent transformation of the available data in the literature. Although ORT is still based on subjective statements from individual authors in the literature on logical relationships between parcellation schemes (the inter-map relationships), it (i) selects clearly described and logically consistent inter-map relationships among thousands of such judgements, (ii) makes all primary data, assumptions and parameters completely transparent, (iii) is published in mathematical detail (Stephan & Kötter 1999; Stephan *et al.* 2000a), and (iv) allows for reproducible

transformations whenever applied to the same data and parameters. ORT thus represents a workable solution to the parcellation problem until coordinate-based experimental data and appropriate spatial mapping techniques are available for primate tracing studies (compare the articles by Mazziotta *et al.*, this issue, and Press *et al.*, this issue).

(iv) *Post-processing of transformation results*

In the current version of CoCoMac, the initial integration of the primary projection data is restricted to redundant or contradictory statements referring to the same brain map (see §2c(ii)). Therefore, overlapping statements still exist after this process of integration—only distributed across several parcellation schemes. Subsequently, each of these statements is converted by ORT from its original description to a common target map chosen by the user. The set of projections that results from this transformation consequently also contains redundant and contradictory information. In contrast to integrated primary data, this is now directly visible since all projections are described on the basis of the same map. Thus, a final step of integration is required.

Basically, this post-transformation integration is performed according to the same principles as the pretransformation integration (see §2c(ii)), i.e. by evaluating two conflicting projections via the associated *PDC_Site* and ECs. As an additional parameter, however, one needs to consider the ‘correctness’ of the mappings along the respective transformation paths. Each transformation path contains a series of inter-map relationships (represented by RCs) each of which may be stated on a different level of precision in the literature and is thus more or less securely interpretable—this is reflected by the *PDC_RC* for each inter-map relationship in a given path. A reasonable criterion to assess the overall reliability of a transformation path is to search the ‘lowest’ *PDC_RC* of the entire path (‘lowest’ according to the user-chosen PDC hierarchy). As a real-life analogy, this procedure is comparable with estimating the durability of an iron chain by assessing the strength of its weakest element.

Briefly summarized, CoCoMac integrates any given set of redundant or contradictory transformed projections by a series of steps that compare: (i) the precision of the identification of source and target sites of the projections (*PDC_SourceSite*, *PDC_TargetSite*), (ii) the precision of inter-map relationships that underlie the transformation (*PDC_RC*), (iii) the absence or existence of the projections and, for the latter case, the quality of existence (*EC_SourceSite*, *EC_TargetSite*) as well as the associated PDCs (*PDC_EC_SourceSite*, *PDC_EC_TargetSite*).

(v) *User-defined parameters of data analysis*

In previous sections, we mentioned that the user can set various parameters to specify in detail how data are initially selected, how primary projections are composed and integrated, how these integrated projections are transformed within CoCoMac, and how transformed data are eventually integrated. This is in accordance with axioms 1, 2, and 5 above (§2a) that postulate that the data from the literature should be represented in their original form in CoCoMac without any implicit interpretation or analysis, and that the user should be able to interpret these primary data from different angles, using different assumptions, thresholds and parcellation schemes.

The number of parameters influencing data analysis in CoCoMac currently amount to more than 20. Before each analysis or transformation, all parameters must be specified by the user; after each analysis, all parameters are saved with the

actual results for documentation. Table 4 summarizes the available parameters and their influence on data analysis in CoCoMac.

(d) **Technical implementation**

Following our initial conception of ORT (Stephan & Kötter 1998), we started developing CoCoMac as a single-user relational database using MS Access 97. In the meantime, CoCoMac has grown into a multi-user relational database on the basis of MS Access 2000 and its replication technology. Organizationally, we are currently using one design master and four replicas distributed across a local network. This set-up allows the individual data collators to work independently and to synchronize their replicas with the design master at regular intervals.

Apart from the data that are currently distributed across 52 tables, CoCoMac includes several modules of code for checking, analysing, as well as transforming data by ORT. These routines currently amount to approximately 7000 lines of Visual Basic for Applications code. As a next step, CoCoMac will be made more easily available to the neuroscientific community via an online internet interface at <http://www.cocomac.org>.

3. RESULTS

(a) **Current contents**

CoCoMac currently contains information on substantial parts of the cortico-cortical network, covering all main regions of the cortex. The degree to which connectivity information is available, however, varies considerably between different cortical regions. While we have already entered into CoCoMac virtually all data on connectivity within the prefrontal cortex, much less information is so far available on auditory or somatosensory cortex. As for subcortical data, CoCoMac currently contains data on thalamic, amygdalar and hippocampal projections, particularly to the prefrontal cortex, but also to other cortical and subcortical regions. At the moment, we extend the subcortical data to visual structures like the lateral geniculate nucleus and superior colliculus.

On 19 December 2000, CoCoMac contained data from 270 papers that together delineated or unspecifically adopted 4723 brain sites (see table 5); 5475 statements in the literature referred to 5395 different inter-map relationships between these brain sites. Concerning the experiment-related data, the papers described 278 different tracing methods that were applied by 1344 injections. These injections yielded descriptions of 10 806 labelled and 9298 unlabelled brain sites, giving a total of 20 104 distinct experimental statements on anatomical projections. With regard to the statements on labelled brain sites, 3515 of them (32.5%) contained additional information on the laminar pattern and 1103 of them (5.5%) also provided quantitative data (e.g. on absolute numbers of labelled somata or terminals).

(b) **Analysing prefrontal connectivity: an example of the practical use of CoCoMac**

As outlined above, CoCoMac allows its users to integrate and transform data according to individual criteria. The resulting datasets can then be passed on to other software, e.g. for models of brain systems based on real connectivity or for statistical analysis of connectivity

Table 5. A quantitative survey of the contents of CoCoMac at 19 December 2000.

data type	name of table	number entries at 19 December 2000
bibliographical	<i>Literature_Main</i>	270
	<i>Literature_Articles</i>	248
	<i>Literature_Books</i>	5
	<i>Literature_BookChapters</i>	17
mapping	<i>BrainMaps</i>	262
	<i>BrainMaps_BrainSiteAcronyms</i>	1510
	<i>BrainMaps_BrainSites</i>	4723
	<i>InterMapRelations</i>	5395
	<i>InterMapRelations_References</i>	5475
	<i>Methods</i>	278
	<i>Methods_Animals</i>	274
	<i>Methods_RadioactiveTracers</i>	104
experiment-related	<i>Injections</i>	1344
	affecting > 1 brain site	515
	<i>Injections_Laminae</i>	109
	<i>LabelledSites_Descriptions</i>	1792
	<i>LabelledSites_Data</i>	20 104
	existing label	10 806
	absent label	9298
	with quantitative data	1103
	<i>LabelledSites_Data_Laminae</i>	3515
	<i>LabelledSites_Data_QD</i>	1103

patterns. Here, we demonstrate an example of the latter approach applied to connectivity between areas of primate prefrontal cortex. Some of the results presented below are also described by Passingham *et al.* (2001).

In a previous analysis of primate prefrontal connectivity (Northoff *et al.* 2000), we had demonstrated that two predefined groups of prefrontal areas, a lateral and an orbitomedial one, possess connectional patterns that are significantly different from each other. Recent analyses re-addressed this question in a more sophisticated way (Kötter *et al.* 2001a; Passingham *et al.* 2001): are there connectivity-defined 'families' of areas that can be determined by multivariate classification techniques, independent of any *a priori* hypothesis? Furthermore, is it necessary that connections are classified by strength, or is knowledge on mere existence or absence of connections sufficient to determine principles of connectional organization in prefrontal cortex?

As of October 2000, CoCoMac contained 127 injections that delivered information on cortico-cortical connections between prefrontal areas. After integration, these primary data were mapped by ORT to a common parcellation scheme, i.e. Walker's (1940) map of prefrontal cortex. Three different data sets were generated by CoCoMac: two sets of graded connections for which higher (HIGH, figure 6a) and weaker (LOW, figure 7a) projection strengths were favoured, respectively, when integrating similarly precise but contradictory reports, and a set of binary connection data which distinguished merely between existing versus absent connections (BIN, figure 8a). Each of these datasets was investigated by hier-

(a)

a)

HIGH

	W10	W11	W12	W13	W14	W24	W25	W45	W46	W8A	W8B	W9
W10		X	3	1	3	X	3	0	0	X	0	X
W11	3		3	3	X	X	X		0	0	0	3
W12	3	3		3	3	1	3	0	X	X	X	X
W13	1	X	3		3	1	3	0	0	0	X	X
W14	3	X	X	3		X	3	0	0	0	0	X
W24	0	3	1	0	X		3	1	1	1	X	3
W25	3	X	3	2	3	3		0	X	0	X	3
W45	1	3	3	1	1	1	0		X	2	X	X
W46	3	2	X	0	0	1	1	2		X	X	X
W8A	X	0	2	0	0	1	0	2	X			X
W8B						3			X	X		X
W9	X	X	X	1	0	3	X	X	X	X	X	

(b)

Figure 6. (a) Matrix of graded connections between prefrontal areas (W) according to Walker (1940). Rows contain efferents and columns contain afferents of the individual areas. Connection strengths are given by 0 (absent), 1 (weak), 2 (medium), 3 (strong) and X (existing connection of unknown strength). Blank fields indicate either lack of information or ambiguous mapping to the chosen target map of Walker. During generation of this data set, higher connection strengths were preferred over weaker ones when contradictory reports of equivalent precision were integrated (HIGH). (b) HCA of Spearman correlation coefficients between the graded connectivity vectors of all prefrontal areas as shown by figure 6a. Using an Euclidean distance metric and complete linkage, the analysis starts on the left (distance = 0), consecutively merging areas or clusters of areas, respectively, at the distance defined between their two most dissimilar areas.

archical cluster analysis (HCA), a multivariate statistical technique that allows assessment of hierarchical relationships of (dis)similarity between connectional patterns of different brain structures. Details of HCA, as well as the parameters used for selection and integration of primary data in ORT, are described in Appendix C.

All analyses showed a clearly clustered configuration of prefrontal connectivity patterns and identified similar

Figure 7. (a) Matrix of graded connections between prefrontal areas (W) according to Walker (1940). The difference from the matrix shown by figure 6a is that during generation of this data set, lower connection strengths were preferred over higher ones when contradictory reports of equivalent precision were integrated (LOW). (b) HCA of Spearman correlation coefficients between the graded connectivity vectors of all prefrontal areas as shown by (a) (see legend to figure 6b).

groupings of areas (figures 6b–8b). Overall, three groups of prefrontal areas are thus distinguishable on grounds of their connectional patterns: (i) lateral areas 8A, 45 and 46, (ii) orbitomedial areas 10, 11, 12, 13, 14 and 25, and (iii) areas 9 and 24. These results were confirmed by multidimensional scaling (MDS) analyses of the same datasets. These independent analyses delivered the same three groups of areas (not shown here, but see Passingham *et al.* (2001) for the MDS results of the HIGH data).

It is remarkable that the HCA results of the two graded datasets (HIGH, LOW) were more different than the results of HIGH and BIN datasets, which showed very similar configurations (compare figures 6b and 8b). The main difference between HIGH and BIN was the

Figure 8. (a) Matrix of binary connections between prefrontal areas (W) according to Walker (1940). Existing connections are designated by 'E', while projections whose absence has explicitly been stated in the literature are represented by 'A' (BIN). (b) HCA applied to 'simple matching' (M) correlation coefficients between binary connectivity vectors of prefrontal areas (see legend to figure 6b).

grouping of area 8B whose interpretation is insecure due to lack of data on its efferents. Furthermore, it is obvious from figures 6b and 8b that although both analyses identified identical groups of areas, HIGH resulted in a better separation of the three groups than BIN. This property is not only visualized by the structure of the cluster tree, but also from the much higher distances at which the three groups are amalgamated in HIGH. While LOW maintained the general distinction between lateral and orbitomedial areas, it merged the orbitomedial cluster with areas 9 and 24 (figure 7b). Also, LOW pointed to two subclusters in the orbitomedial group (areas 11, 14 and 25 versus 10, 12 and 13). Finally, as one would expect, the analyses also found slightly different hierarchical relationships within the three groups. For example, while all analyses demonstrated high similarities between areas 11, 14 and 25, the exact relationships varied.

This analysis can easily be varied by changing further parameters of data integration and transformation in CoCoMac: one can choose different maps according to which the results are described, extract data on laminar patterns of the projections and include these into subsequent analyses, and so forth. Even this simple example demonstrates that the flexibility to generate different types of datasets depending on the user-chosen parameters has at least two major advantages: (i) it allows the investigation of the same brain system from different perspectives and (ii) it controls for the influence of certain assumptions or chosen parcellation schemes on the results of a given analysis.

4. DISCUSSION

In this article, we have described the methodological strategy of CoCoMac, a relational database of the anatomical brain connectivity in the macaque monkey brain. While we did not intend to describe the analytical components of CoCoMac in full mathematical detail, this paper presents the general theoretical concepts underlying our approach to representing and analysing neuro-anatomical data of considerable complexity, variety and uncertainty.

We are aware that the approach presented here is not the only possible one that can successfully deal with the methodological problems of connectivity data. Our concept of data representation abstracts from spatial information and treats brain sites and brain maps as conceptual entities in a multidimensional feature space. This basic philosophy accounts for both the strength and limitations of our approach. The main advantages of our method have been described above: strict separation of data representation and analysis, maintenance of original parcellation and nomenclature with a minimum of interpretation, flexible transformation to user-chosen brain maps, flexible analysis according to user-chosen parameters and transparency of data representation and analysis. Limitations of CoCoMac result mainly from the fact that spatial information is only considered in an abstract sense by the ECs and RCs. Therefore, and in contrast to connectivity databases that use spatial reference systems (e.g. Press *et al.*, this issue), users of CoCoMac get only limited information on the topography of injections and label or on the exact spatial relationship between brain sites from different maps. While this spatial information is not crucial for integrating many aspects of connectivity across large datasets, it may still be essential for questions regarding individual projections. Whether CoCoMac is the tool of choice for look-up and analysis of connectivity data depends on the individual user's intentions.

The current state of implementation of CoCoMac can be described as a 'functional construction site'. That is, although the design of the database has been completed and many tools of analysis have been programmed, we are continuously improving and extending the current implementation. For example, CoCoMac still lacks a comprehensive and user-friendly interface, with the result that only experts can currently make use of its functionality. Also, the performance of several key algorithms needs to be improved and some analytical features only

exist in provisional form so far. CoCoMac is in a state of on-going evolution so that its current implementation is only one among different possibilities.

With respect to data representation, CoCoMac represents data on the basis of brain sites that are delineated within various parcellation schemes. The *BrainSite* objects that are currently defined in CoCoMac for this purpose can represent various modules of brain organization (see table 2). Another modern database, NeuroScholar (see Burns, this issue), is based on a similar, still more sophisticated concept. NeuroScholar contains two basic classes of objects for representation of data: the *brainVolume* class and the *neuronPopulation* class. Conceptually, *brainVolume* objects in NeuroScholar are very similar to *BrainSite* objects within CoCoMac. Indeed, NeuroScholar has adopted several properties of *BrainSite* objects in CoCoMac (e.g. ECs) and will also incorporate ORT routines for mapping data based on *brainVolume* objects between different parcellation schemes. The *neuronPopulation* class represents a population of abstract neurons (including a population of one element, i.e. a single abstract neuron) with neuronal subcomponents like somata, axons and dendrites. NeuroScholar thus also allows for very precise representations of data from single-neuron tracing studies as well as from studies using multiple tracers. In contrast, the *BrainSite* objects currently implemented in CoCoMac allow the resolution of single neuronal somata but do not yet include representations for subcellular components of neurons. Due to the modular design of CoCoMac, it is easily feasible, however, to add suitable objects for this purpose. Most likely, the two databases will continue to exchange methodological advances in the future. This development is supported by efforts towards creating conceptual and technical interfaces that will allow a mutual exchange of data between NeuroScholar and CoCoMac (Burns *et al.* 2001).

This specific comparison between CoCoMac and NeuroScholar leads to the question whether strategies of data representation can be classified in a more general way. We suggest that there are five main dimensions that altogether characterize the overall mode of data representation for a specific database: (i) reference systems, (ii) transformation, (iii) coding, (iv) independence, and (v) uniqueness of data.

- (i) Reference systems. Each experimental finding needs a precise description of the part of the nervous system it relates to, i.e. a reference system within the brain. The definition of such references can be achieved either by 3D coordinates or on the basis of parcellation schemes that contain descriptive units (including cortical areas, nuclei, laminae, columns, or even single neurons). Spatial coordinates (e.g. within a stereotactic reference system) provide a precise reference for an individual brain and are indispensable for neurosurgery and for *in vivo* imaging studies, for instance by means of the Talairach system (Talairach & Tournoux 1988). Because of the limited information available for an individual brain and because of the large interindividual variability in brain morphology, however, coordinates alone are not the final solution for representing brain data. A promising approach to

interindividual variability is the application of magnetic resonance imaging-based nonlinear warping methods for matching individual brains to the template of a reference brain (Ashburner & Friston 1999; Roland & Zilles 1996; Schormann & Zilles 1998). As for primate tracing studies, however, neither spatial coordinates provided in the description of connectivity nor sophisticated warping procedures are readily available yet. Instead, the standard procedure is to relate the experimental results to the components of one of the many parcellation schemes that have been suggested for the primate brain. A difficult problem of this approach, however, results from the large variety of existing parcellation schemes and the abundant incompatibility between them (see § 2b(ii)).

- (ii) Transformation of data. If parcellation schemes are used as a reference system for brain data, a general decision must be made whether the database entries preserve the nomenclature for brain locations as used by the original article or not. Choosing non-transformed locations for brain data means that the database represents the locations in the same way as in the original article (i.e. by adopting the nomenclature and parcellation scheme of the original report). Preferring transformed locations for brain data means that the locations as stated by the original report are transformed to some 'reference template' (e.g. to a 'reference parcellation scheme' that was defined by the data collator before the database construction was started).
- (iii) Coding of data. Similar to the above question of data transformation, each database requires the decision whether the various properties of the primary data from the literature should be coded or not. Non-coded data are, for example, citations of paragraphs from the text, numeric values from tables (e.g. absolute numbers of labelled cells), or scanned figures. Examples for coded data are our classifications for RCs, ECs, density, laminar patterns or PDCs. Another classical coding scheme for data is a classification for laminar patterns of projections that does not report the individual source and target laminae of a given projection, but classifies the entire pattern as ascending, lateral or descending (Felleman & Van Essen 1991).
- (iv) Independence of data. Very often, the literature contains multiple statements (spread out within one or across several articles) that relate to the same question, e.g. whether a projection exists from structure A to structure B. Such multiple statements can either be compatible or contradictory. When representing these statements in a database, one option is to condense all currently known statements related to the same question into a single entry of the database (summarized data representations) and list the references for the individual statements that underlie this entry (e.g. Felleman & Van Essen 1991; Scannell *et al.* 1995). Alternatively, for each individual statement in the literature one can create an individual representation in the database based on the original description (independent data representations). Expressed more formally, if there are N statements

in the literature on the existence or properties of a given projection (regardless of whether these are (i) based on the same or a different nomenclature, (ii) compatible or contradictory, and (iii) found in one or in different articles), our definition of independent data representations demands that the database contains N complete and entirely independent entries that exactly reflect the N statements in the literature.

- (v) Uniqueness of data. A fifth distinction concerns the question whether each datum is entered only once (unique representation) or possibly several times by different observers (multiple representations). This question also concerns annotations to the data of a database. For example, a particular database may contain single representations of each primary datum, but allow for multiple annotations of different authors to each datum.

Table 6 lists the constellations of these five dimensions as implemented by six published connectivity databases. It is obvious from this table that all connectivity databases prior to NeuroScholar and CoCoMac pursued a relatively uniform mode of data representation: unique representations of summarized and coded data described by brain maps that were transformed to a predefined reference parcellation scheme. Considering the complex structure of connectivity data and the large variability of underlying brain maps, this combination is a time-saving and pragmatic strategy. It has become clear, however, that pretransformation to a reference map and the use of summary data is problematic in the long run. These earlier data collations require labour-extensive re-evaluations as the evolution of parcellation schemes continues and as errors need to be corrected. Drawing from this experience, one of the main goals of the NeuroScholar and CoCoMac projects was to introduce flexible and effective methods for using non-transformed and independent data. An additional focus of NeuroScholar is quality control by keeping multiple representations that reflect different interpretations of the same data by different scientists (see Burns, this issue). It should be acknowledged that the methodical advances of NeuroScholar and CoCoMac have benefited from the easy availability of powerful database software that was not equally obtainable at the time when the first connectivity databases were created.

Evaluating the conceptual design of CoCoMac and its current implementation, we are confident that this database fulfils the goals and axioms that we defined before starting the construction. In its current version, CoCoMac is a fully workable database that covers a considerable part of the macaque brain (see § 3) and allows for flexible, user-controlled extraction, integration, analysis and transformation of connectivity data. Its data and tools have already proven to be useful for both theoretical and for experimental studies (Bozkurt *et al.* 2001; Kötter *et al.* 2001a,b; Hautzel *et al.* 2000; Northoff *et al.* 2000; Passingham *et al.* 2001). In the future, we intend to make CoCoMac publicly available on the internet (<http://www.cocomac.org>) so that users can look up individual data, perform online queries, and download data analysed offline. The capability to perform online

Table 6. Schematic overview on strategies of data representation used by connectivity databases.

(A slash (/) indicates that two approaches were used in combination.)

	database					
	Felleman & Van Essen (1991)	Young (1993)	Scannell <i>et al.</i> (1995, 1999)	Burns (1997)	Neuro- Scholar (Burns 1999, 2000)	CoCoMac (this report)
coordinates (C) versus parcellations (P)	P	P	P	P	P	P
transformed (T) versus non-transformed (nT) data	T	T	T	T	nT	nT
coded (C) versus non-coded (nC) data	C	C	C	nC/C	nC/C	nC/C
summarized (S) versus independent (I) data	S	S	S	S/I	I	I
unique (U) versus multiple (M) representations	U	U	U	U	M	U/M

analyses and transformations for large datasets, however, will be limited. This results from the fact that some of the algorithms in CoCoMac have a minimal run-time complexity above the order of n^2 (more precisely, the construction of the transformation graph has a minimal complexity of $O(n^2 \cdot \log n)$, Güting 1992) and thus time for computation increases exponentially as new brain sites are added. Alternatives include the regular updating of a series of 'standard' analyses and transformations for offline download, or the division of the database into computationally tractable subsets of data that can be dealt with in real time by dedicated processors.

While most of the current demand for primate connectivity data comes from neuroscientists who require these data to inform other domains of research (notably the interpretation of functional imaging data), we also hope to stimulate direct feedback from the fibre tracing community. For example, as an external control in addition to our internal proofreading, it would be helpful if the data contained in CoCoMac could be cross-checked and enhanced by the authors who produced the data. In the long term, fuller publication of experimental data and formalized procedures for data upload into such databases may be established. The systematic upload of data during the course of publication has been a major reason for the power and success of databases in molecular biology. It shortens the arduous process of entering data from the literature enormously and guarantees a better documentation of the experimental data. There are good arguments to include even unpublished data since these can bear highly valuable information and would otherwise be entirely lost to the community (Maurer *et al.* 2000).

Another important goal in the future will be to establish internet interfaces with other neuroscientific databases and with software tools that can be used for online analysis of connectivity data. The link to other databases will enable users to complement connectivity data from CoCoMac with other data modalities (e.g. on transmitter receptors—see Kötter *et al.* (2001b)). Also, it will be important to compare brain connectivity between different species. As a first step, we intend to pioneer the data exchange between NeuroScholar and CoCoMac. This is facilitated by the similar design of the two databases with regard to data representation (see § 2b,4).

As a final note, the enormous progress in the fields of diffusion tensor imaging (DTI) (e.g. Conturo *et al.* 1999)

and *in vivo* magnetic resonance tracing (Pautler *et al.* 1998) should be emphasised. These rapidly developing methods prompt hopes that, in the future, *in vivo* human connectivity data could be obtained non-invasively. The most promising approach is to apply fibre-tracking algorithms to high resolution DTI data (Poupon *et al.* 2000). Human connectivity data would be of enormous scientific and clinical significance, but the specificity and sensitivity of DTI tracking methods are still unclear. Unfortunately, fine-grained anatomical connectivity data are not available for man, so we lack a standard with which DTI results could be compared. Therefore, testing fibre-tracking methods on animal DTI data will be indispensable to assess the validity of these methods. In conclusion, powerful databases of primate connectivity do not only help to address fundamental questions of brain organization, but will also be of major importance for the evaluation of methods related to measuring human connectivity *in vivo*.

We are indebted to C. C. Hilgetag, Boston, J. W. Scannell, Newcastle, A. Schleicher, Düsseldorf, F. T. Sommer, Ulm, U. Stockhorst, Düsseldorf and K. Zilles, Düsseldorf for highly valuable discussions throughout the development of CoCoMac. We also thank two reviewers for constructive comments that helped to improve this article. This work was supported by the German National Merit Foundation (A.B., K.E.S.), the Neuroscientific Graduate School (Neuro-Graduiertenkolleg Pathologische Prozesse des Nervensystems) at the University of Düsseldorf (K.E.S.), a Wellcome Trust Collaboration Grant (M.P.Y. and R.K.), and the German Research Council (R.K., LIS 4-554 95 (2) Düsseldorf).

APPENDIX A

One basic mode of information within CoCoMac is the extent of label or injected tracer substance within a given brain site (see definition of ECs in § 2c(ii)). The various types of brain sites in CoCoMac differ in conceptualizing this kind of information either in 3D space or as a projection onto a two-dimensional plane (see table 2, third column). *BS_IsoCtx_2D*, for example, equals an isocortical area which is seen from above, i.e. where all information on labelled neurons is projected from three-dimensional space onto a 2D sheet. *BS_IsoCtx_3D*, on the other hand, is conceptualized as a truly 3D structure. Since there is a clear and universally applied convention for the lamination of isocortical areas (i.e. six layers that are numbered

ascendingly from the outside to the inside), labelling of individual laminae can be represented by intrinsic properties of *BS_IsoCtx_2D/3D* representation objects (see table 2, fourth column, and §2b(iii)). This approach considerably simplifies data representation for isocortical areas, but is no longer feasible for those laminated brain regions in which the literature disagrees about nomenclature and numbers of layers. This is the situation for both allocortical and layered subcortical structures (e.g. lateral geniculate nucleus and superior colliculus). In these cases, to avoid the considerable confusion that would arise from a strict coding scheme spanning across such incompatible systems of lamination, each individual lamina is represented as a single brain site in CoCoMac by using the *BS_Lamina_Ctx_2D/3D* and *BS_Lamina_SubCtx_3D* representation objects, respectively, that are then linked to representation objects for the brain area (i.e. *BS_AlloCtx_2D/3D* or *BS_Nucl_SubCtx_3D* representation objects).

Dealing with allocortical areas represented by *BS_AlloCtx_2D/3D* objects, the representation of data on single laminae by independent representation objects (i.e. *BS_Lamina_Ctx_2D/3D*) is a flexible and precise method, yet it introduces the problem of mapping data from the laminae to the cortical area that they belong to. The RCs, as defined in §2a, are working equally well within either two or three dimensions, but cannot be used to relate two- or three-dimensionally conceptualized brain sites to each other or to serve as a basis for projecting the ECs of several two- or three-dimensionally defined brain sites onto one common plane. Therefore, since the initial description of RCs and their use within ORT (Stephan & Kötter 1998, 1999; Stephan *et al.* 2000a), we have recently extended ORT by adding further RCs that define interfaces between two- and three-dimensionally defined objects. Adding these new RCs requires associated extensions to the algebra of transformation and the graph-theoretical procedures within ORT whose descriptions will be the subject of a forthcoming article.

APPENDIX B

As described in section §2b(iii), laminar patterns—both of injected tracer substance and transported label—are described differently in CoCoMac depending on whether the concerned brain site is a typical six-layered isocortical area or shows a different laminar structure. In case of isocortical areas, the laminar pattern is represented by a sequence of six alphanumeric place-holders: 0: lamina contains no label; 1: lamina is weakly or sparsely labelled in comparison with other laminae of the same area; 2: lamina is moderately labelled in comparison with other laminae of the same area; 3: lamina is densely or heavily labelled in comparison with other laminae of the same area; X: lamina is labelled (however, there are no data on the relative intensity compared with other laminae of the same area); ?: information about potential labelling of this lamina is ambiguous or missing, therefore it remains unclear whether this lamina contains label or not.

For example, the original description ‘...terminals were labelled over layer I, fewer were labelled in layers II and III, and the fewest were labelled in layer IV’ (Vogt

& Pandya 1987, p. 281) of label in the six-layered cingulate cortex would be represented in CoCoMac by the string ‘322100’. It should be noted, however, that due to the often rather coarse descriptions of relative label strengths between different laminae in the literature and the merely ordinal character of the numeric place-holders, our coding scheme sometimes allows for several equivalent representations. For example, if the statement ‘...the labelled cells were located in layer VI and, to a lesser extent, layer V...’ (Carmichael & Price 1995a, p.628) is represented in CoCoMac, both ‘000012’ and ‘000023’ would adequately express the ranking of labelled laminae. It is important to keep in mind that the numeric values in this scheme do not represent absolute values of strength, but merely indicate the relative strength of label compared with other laminae of the same brain site.

The use of this alphanumeric coding scheme is dependent on a basal and generally shared concept of isocortical lamination, i.e. that isocortical areas have six layers, which are numbered in an ascending manner from the outermost layer to the innermost layer. In contrast to isocortical areas, such a general notion does not exist for allocortical areas or those subcortical structures that also show a layered structure (e.g. lateral geniculate nucleus and superior colliculus). Instead, different authors vary in terminology and even number of identified layers in these structures (for a review on this topic in subcortical visual structures, see Kaas & Huerta (1988)). Therefore, already at the early stage of defining brain sites in CoCoMac, each of the laminae of allocortical areas and layered subcortical nuclei are represented as individual brain sites in CoCoMac, using the appropriate *BrainSite* types (see Appendix A). When subsequently representing label in CoCoMac that was experimentally found in these structures, each of the affected laminae as defined by the author is listed separately.

APPENDIX C

When computing the three data sets shown by figures 6a, 7a and 8a, the following parameters were used for selection of primary data and for data integration in CoCoMac (compare §2c(v) and table 4): target map for transformation is Walker’s (1940) map of prefrontal cortex; PDC hierarchy for identification of brain sites = A > C > F > D > H > L > I > M > B > E > G > J > N > K > O > P > Q (compare figure 2 and §2b(v)); the remaining four PDC hierarchies which are not described in detail in this paper have been defined accordingly); no PDC thresholds; ‘strict’ interpretation of injections affecting more than one brain site (STRICT, see §2c(ii)); rejection of injections with all other methodical problems; only ipsilateral projections; both right and left hemispheres; all tracer substances; and both genders of all macaques species.

The parameter that was varied between the three data-sets concerned the way in which connection strengths were integrated across reports. For the HIGH data, higher projection strengths were preferred over lower ones when integrating findings of similarly precise description but with differently reported projection strength. For the LOW data, the reverse rule was followed. In case of the BIN data, projection strengths

were completely disregarded and only the existence or absence of projections were considered.

After the connectivity data had been extracted from CoCoMac, they were analysed by means of the software package SYSTAT (v. 9, SPSS Inc.). The first step was a correlation analysis to obtain comparable indices of (dis)similarity between the connectivity vectors of all individual areas (note that the term 'connectivity vector' as used here refers to the concatenation of all efferents and afferents for a given area). For graded connectivity vectors (HIGH or LOW), Spearman correlation coefficients were computed. Binary connectivity vectors (BIN) were cross-correlated using the 'simple matching' dichotomy (also known as 'M coefficient', see Backhaus *et al.* (1996)). The connectivity vectors of all areas were thus transformed into feature vectors composed of correlation coefficients, and the three connectivity matrices were thus characterized by symmetric matrices of correlation coefficients. These correlation matrices were then analysed by a multivariate statistical technique, Hierarchical Cluster Analysis (HCA). In the context of our analysis, HCA reveals hierarchical relationships of (dis)similarity between areas with regard to their connectional patterns. Briefly summarized, the general principle of HCA is to iteratively determine all pairwise distances between the feature vectors of all considered elements. Starting with all elements as individual entities, at each step the two least distant elements are merged. Depending on the scale of the data, different distance metrics are available. Since we apply HCA to vectors of correlation coefficients, a Euclidean distance metric was appropriate for our analysis. As merging criterion, we used 'complete linkage' for which merging of clusters occurs at the distance defined by the two most distant elements (i.e. with the two most dissimilar feature vectors) of the two clusters.

REFERENCES

- Aggleton, J. P. 1986 A description of the amygdalo-hippocampal interconnections in the macaque monkey. *Exp. Brain Res.* **64**, 515–526.
- Aggleton, J. P. & Mishkin, M. 1984 Projections to the thalamus in the cynomolgus monkey. *J. Comp. Neurol.* **222**, 56–68.
- Ashburner, J. & Friston, K. J. 1999 Nonlinear spatial normalization using basis functions. *Hum. Brain. Mapp.* **7**, 254–266.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. 1996 *Multivariate Analysemethoden*. Berlin: Springer.
- Barbas, H. & Pandya, D. N. 1987 Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J. Comp. Neurol.* **256**, 211–228.
- Barbas, H. & Pandya, D. N. 1989 Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* **286**, 353–375.
- Barone, P., Batardiere, A., Knoblauch, K. & Kennedy, H. 2000 Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* **20**, 3263–3281.
- Bates, J. F. & Goldman-Rakic, P. S. 1993 Prefrontal connections of medial motor areas in the rhesus monkey. *J. Comp. Neurol.* **336**, 211–228.
- Bortz, J. & Döring, N. 1995 *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Bozkurt, A., Kamper, L., Stephan, K. E. & Kötter, R. 2001 Analysis of amygdalo-thalamo-prefrontal connectivity: structure–function relationships in associative learning. *Neurocomputing* **38–40** (Special Triple Volume), 1135–1140.
- Brodmann, K. 1905 Beiträge zur histologischen Lokalisation der Grosshirnrinde. III, Die Rindenfelder der niederen Affen. *J. Psychol. Neurol.* **4**, 177–226.
- Brodmann, K. 1909 *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig, Germany: Barth.
- Burns, G. A. P. C. 1997 Neural connectivity of the rat: theory, methods and applications. DPhil thesis, University of Oxford, UK.
- Burns, G. A. P. C. 1999 Neuroscholar 1.00, a neuroinformatics databasing website. *Neurocomputing* **26–27**, 963–970.
- Burns, G. A. P. C. 2001 Knowledge mechanics and the NeuroScholar project, a new approach to neuroscientific theory. In *Computing the brain: a guide to neuroinformatics* (ed. M. A. Arbib & J. Grethe), pp. 319–335. San Diego, CA: Academic Press. (In the press.)
- Burns, G. A. P. C. & Young, M. P. 2000 Analysis of the connectional organisation of neural systems associated with the hippocampus in rats. *Phil. Trans. R. Soc. Lond. B* **355**, 55–70.
- Burns, G. A. P. C., Stephan, K. E., Ludaescher, B., Gupta, A. & Kötter, R. 2001 Towards a federated neuroscientific knowledge management system using brain atlases. *Neurocomputing* **38–40** (Special Triple Volume), 1633–1641.
- Carmichael, S. T. & Price, J. L. 1994 Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* **346**, 366–402.
- Carmichael, S. T. & Price, J. L. 1995a Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J. Comp. Neurol.* **363**, 615–641.
- Carmichael, S. T. & Price, J. L. 1995b Sensory and premotor connections of the orbital and medial prefrontal cortex. *J. Comp. Neurol.* **363**, 642–664.
- Carmichael, S. T. & Price, J. L. 1996 Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J. Comp. Neurol.* **371**, 179–207.
- Codd, E. F. 1970 A relational model for large shared data banks. *Comm. ACM* **13**, 377–387.
- Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H. & Raichle, M. E. 1999 Tracking neuronal fiber pathways in the living human brain. *Proc. Natl Acad. Sci. USA* **96**, 10 422–10 427.
- Crosby, E. C. & Humphrey, T. 1941 Studies of the vertebrate telencephalon. II. The nuclear pattern of the anterior olfactory nucleus tuberculum olfactorium and the amygdaloid complex in adult men. *J. Comp. Neurol.* **74**, 309–352.
- Felleman, D. J. & Van Essen, D. C. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47.
- Floyd, R. W. 1962 Algorithm 97: shortest paths. *Comm. ACM* **5**, 345.
- Gahr, M. 1997 How should brain nuclei be delineated? Consequences for developmental mechanisms and for correlations of area size, neuron numbers and functions of brain nuclei. *Trends Neurosci.* **20**, 58–62.
- Graepel, T. & Obermayer, K. 1999 A stochastic self-organizing map for proximity data. *Neural Comput.* **11**, 139–155.
- Guilford, J. P. 1954 *Psychometric methods*. New York: McGraw Hill.
- Gütting, R. H. 1992 *Datenstrukturen und Algorithmen*. Stuttgart: Teuber.
- Hautzel, H., Mottaghy, F. M., Horwitz, B., Dammers, J., Beu, M., Kötter, R., Stephan, K. E., Müller-Gärtner, H. W. & Krause, B. J. 2000 Network analysis on verbal working memory: a combined PET and MEG study. *Soc. Neurosci. Abstr.* **26**, 263.7.
- Hilgetag, C. C., O'Neill, M. A. & Young, M. P. 1996 Indeterminate organization of the visual system. *Science* **271**, 776–777.

- Hilgetag, C. C., Burns, G. A. P. C., O'Neill, M. A., Scannell, J. W. & Young, M. P. 2000a Anatomical connectivity defines the organisation of clusters of cortical areas in macaque monkey and cat. *Phil. Trans. R. Soc. Lond. B* **355**, 91–110.
- Hilgetag, C. C., O'Neill, M. A. & Young, M. P. 2000b Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Phil. Trans. R. Soc. Lond. B* **355**, 71–89.
- Jimenez-Castellanos, J. 1949 The amygdaloid complex in monkey studied by reconstructional methods. *J. Comp. Neurol.* **91**, 507–526.
- Jouve, B., Rosenstiehl, P. & Imbert, M. 1998 A mathematical approach to the connectivity between the cortical visual areas of the macaque monkey. *Cereb. Cortex* **8**, 28–39.
- Kaas, J. H. & Huerta, M. F. 1988 The subcortical visual system of primates. In *Comparative primate biology* (ed. J. Erwin), pp. 327–391. New York: Liss.
- Köbber, C., Apps, R., Bechmann, I., Lanciego, J. L., Mey, J. & Thanos, S. 2000 Current concepts in neuroanatomical tracing. *Prog. Neurobiol.* **62**, 327–351.
- Kötter, R. 1999 Trends in European computational neuroscience. *Rev. Neurosci.* **10**, 175–179.
- Kötter, R. & Sommer, F. T. 2000 Global relationship between anatomical connectivity and activity propagation in the cerebral cortex. *Phil. Trans. R. Soc. Lond. B* **355**, 127–134.
- Kötter, R. & Stephan, K. E. 2000 Area interrelations in primate prefrontal cortex. *NeuroImage* **11**, S86.
- Kötter, R., Hilgetag, C. C. & Stephan, K. E. 2001a Connectional characteristics of areas in Walker's map of primate prefrontal cortex. *Neurocomputing* **38–40**, 741–746.
- Kötter, R., Stephan, K. E., Palomero-Gallagher, N., Geyer, S., Schleicher, A., Zilles, K. 2001b Multimodal characterisation of cortical areas by multivariate analyses of receptor binding and connectivity data. *Anat. Embryol.* (In the press.)
- Lienert, G. A. & Raatz, U. 1994 *Testaufbau und Testanalyse*. Weinheim, Germany: Psychologie Verlagsunion.
- Luppino, G., Matelli, M., Camarda, R. & Rizzolatti, G. 1993 Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *J. Comp. Neurol.* **338**, 114–140.
- MacNeil, M. A., Lomber, S. G. & Payne, B. R. 1997 Thalamic and cortical projections to middle suprasylvian cortex of cats: constancy and variation. *Exp. Brain Res.* **114**, 24–32.
- Matelli, M., Camarda, R., Glickstein, M. & Rizzolatti, G. 1986 Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.* **251**, 281–298.
- Matelli, M., Luppino, G. & Rizzolatti, G. 1991 Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. *J. Comp. Neurol.* **311**, 445–462.
- Maurer, S. M., Firestone, R. B. & Sriver, C. R. 2000 Science's neglected legacy. *Nature* **405**, 117–120.
- McGuire, P. K., Bates, J. F. & Goldman-Rakic, P. S. 1991 Interhemispheric integration: I. Symmetry and convergence of the corticocortical connections of the left and the right principal sulcus (PS) and the left and the right supplementary motor area (SMA) in the rhesus monkey. *Cereb. Cortex* **1**, 390–407.
- Merabet, L., Desautels, A., Minville, K. & Casanova, C. 1998 Motion integration in a thalamic visual nucleus. *Nature* **396**, 265–268.
- Mitz, A. R. & Wise, S. P. 1987 The somatotopic organization of the supplementary motor area: intracortical microstimulation mapping. *J. Neurosci.* **7**, 1010–1021.
- Northoff, G. (and 15 others) 2000 Functional dissociation between medial and lateral orbitofrontal cortical spatiotemporal activation in negative and positive emotions: a combined fMRI/MEG study. *Cereb. Cortex* **10**, 93–107.
- Pandya, D. N. & Seltzer, B. 1982 Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey. *J. Comp. Neurol.* **204**, 196–210.
- Passingham, R. E., Stephan, K. E. & Kötter, R. 2001 The anatomical basis of functional localization in the cortex. (Submitted.)
- Pautler, R. G., Silva, A. C. & Koretsky, A. P. 1998 *In vivo* neuronal tract tracing using manganese-enhanced magnetic resonance imaging. *Magn. Reson. Med.* **40**, 740–748.
- Petrides, M. & Pandya, D. N. 1994 Comparative architectonic analysis of the human and the macaque frontal cortex. In *Handbook of neuropsychology* (ed. F. Boller & J. Grafman), pp. 17–58. Amsterdam: Elsevier.
- Poupon, C., Clark, C. A., Frouin, V., Regis, J., Bloch, I., Le Bihan, D. & Mangin, J.-F. 2000 Regularization of diffused-based direction maps for the tracking of brain white matter fascicles. *NeuroImage* **12**, 184–195.
- Powell, T. P. S. & Mountcastle, V. B. 1955 The cytoarchitecture of the postcentral gyrus of the monkey *Macaca mulatta*. *Bull. Johns Hopkins Hosp.* **105**, 108–132.
- Preuss, T. M. & Goldman-Rakic, P. S. 1991 Architectonics of the parietal and temporal association cortex in the strepsirrhine primate *Galago* compared to the anthropoid primate *Macaca*. *J. Comp. Neurol.* **310**, 475–506.
- 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.
- Roland, P. E. & Zilles, K. 1996 The developing European Computerized Human Brain Database for all imaging modalities. *NeuroImage* **4**, 39–47.
- Scannell, J. W. 1997 Determining cortical landscapes. *Nature* **386**, 452.
- Scannell, J. W., Blakemore, C. & Young, M. P. 1995 Analysis of connectivity in the cat cerebral cortex. *J. Neurosci.* **15**, 1463–1483.
- Scannell, J. W., Sengpiel, F., Tovee, M. J., Benson, P. J., Blakemore, C. & Young, M. P. 1996 Visual motion processing in the anterior ectosylvian sulcus of the cat. *J. Neurophysiol.* **76**, 895–907.
- Scannell, J. W., Burns, G. A. P. C., Hilgetag, C. C., O'Neill, M. A. & Young, M. P. 1999 The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex* **9**, 277–299.
- Schormann, T. & Zilles, K. 1998 Three-dimensional linear and nonlinear transformations: an integration of light microscopical and MRI data. *Hum. Brain Mapp.* **6**, 339–347.
- Sporns, O., Tononi, G. & Edelman, G. M. 2000 Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* **10**, 127–141.
- Stephan, K. E. & Kötter, R. 1998 A formal approach to the translation of cortical maps. In *Neural circuits and networks* (ed. V. Torre & J. Nicholls), pp. 205–226. Berlin: Springer.
- Stephan, K. E. & Kötter, R. 1999 One cortex—many maps: an introduction to coordinate-independent mapping by objective relational transformation (ORT). *Neurocomputing* **26–27**, 1049–1054.
- Stephan, K. E., Zilles, K. & Kötter, R. 2000a Coordinate-independent mapping of structural and functional cortical data by objective relational transformation (ORT). *Phil. Trans. R. Soc. Lond. B* **355**, 37–54.
- Stephan, K. E., Hilgetag, C. C., Burns, G. A. P. C., O'Neill, M. A., Young, M. P. & Kötter, R. 2000b Computational analysis of functional connectivity between areas of primate cerebral cortex. *Phil. Trans. R. Soc. Lond. B* **355**, 111–126.
- Talairach, J. & Tournoux, P. 1988 *Co-planar stereotaxic atlas of the human brain*. Stuttgart, Germany: Thieme.

- Ullman, J. D. & Widom, J. 1997 *A first course in database systems*. Upper Saddle River, NJ: Prentice-Hall.
- 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, MA: MIT Press.
- Vogt, B. A. & Pandya, D. N. 1987 Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* **262**, 271–289.
- Vogt, B. A., Pandya, D. N. & Rosene, D. L. 1987 Cingulate cortex of the rhesus monkey. I. Cytoarchitecture and thalamic afferents. *J. Comp. Neurol.* **262**, 256–270.
- Vogt, C. & Vogt, O. 1919 Ergebnisse unserer Hirnforschung. Vierte Mitteilung: die Physiologische Bedeutung der Architektonischen Rindenreizungen. *J. Psychol. Neurol.* **25**, 279–461.
- Von Bonin, G. & Bailey, P. 1947 *The neocortex of Macaca mulatta*. Urbana, IL: University of Illinois Press.
- Walker, E. A. 1940 A cytoarchitectural study of the prefrontal area of macaque monkey. *J. Comp. Neurol.* **73**, 59–86.
- Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442.
- Young, M. P. 1992 Objective analysis of the topological organization of the primate cortical visual system. *Nature* **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. 2000 Brain structure–function relationships: advances from neuroinformatics. *Phil. Trans. R. Soc. Lond. B* **355**, 3–6.
- Young, M. P., Scannell, J. W., O’Neill, M. A., Hilgetag, C. C., Burns, G. & Blakemore, C. 1995 Non-metric multi-dimensional scaling in the analysis of neuroanatomical connection data and the organization of the primate cortical visual system. *Phil. Trans. R. Soc. Lond. B* **348**, 281–308.
- Young, M. P., Hilgetag, C. C. & Scannell, J. W. 2000 On imputing function to structure from the behavioural effects of brain lesions. *Phil. Trans. R. Soc. Lond. B* **355**, 147–161.