THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0907

A graphical anatomical database of neural connectivity

William A. Press[1], **Bruno A. Olshausen**[2*] **and David C. Van Essen**[3]

[1]*Department of Psychology, Stanford University, Stanford, CA 94305, USA*
[2]*Center for Neuroscience (and Department of Psychology), 1544 Newton Court, UC Davis, Davis, CA 95616, USA*
[3]*Department of Anatomy and Neurobiology, Washington University School of Medicine, St Louis, MO 63110, USA*

We describe a graphical anatomical database program, called XANAT (so named because it was developed under the X window system in UNIX), that allows the results of numerous studies on neuroanatomical connections to be stored, compared and analysed in a standardized format. Data are entered into the database by drawing injection and label sites from a particular tracer study directly onto canonical representations of the neuroanatomical structures of interest, along with providing descriptive text information. Searches may then be performed on the data by querying the database graphically, for example by specifying a region of interest within the brain for which connectivity information is desired, or via text information, such as keywords describing a particular brain region, or an author name or reference. Analyses may also be performed by accumulating data across multiple studies and displaying a colour-coded map that graphically represents the total evidence for connectivity between regions. Thus, data may be studied and compared free of areal boundaries (which often vary from one laboratory to the next), and instead with respect to standard landmarks, such as the position relative to well-known neuroanatomical substrates or stereotaxic coordinates. If desired, areal boundaries may also be defined by the user to facilitate the interpretation of results. We demonstrate the application of the database to the analysis of pulvinar–cortical connections in the macaque monkey, for which the results of over 120 neuroanatomical experiments were entered into the database. We show how these techniques can be used to elucidate connectivity trends and patterns that may otherwise go unnoticed.

Keywords: graphical database; neuroanatomy; cortex; thalamus; inference

1. INTRODUCTION

The application of modern pathway tracing techniques over the past several decades has revealed a wealth of information on the connections between brain regions in a variety of different species. Although our current store of knowledge could be considered vast, most of the data exists scattered through journal articles in the form of photographs or diagrams of tissue cross-sections that have been marked (or 'scored') in the locations where labelling was observed after a neuroanatomical tracer was injected in a particular region of the brain. Those sites where labelling was observed provide evidence that a connection exists between them and the location where the injection was made. Currently, few methods exist for viewing this wealth of data in a unified format that accurately summarizes the existing state of our knowledge. As the amount of neuroanatomical data increases, it will become increasingly difficult to access and assimilate this information using conventional literature searches and human memory. Just as genome and protein databases have proven critical for molecular biologists, neuroanatomical databases are now becoming an important tool for organizing the increasing amount of information on neural connection pathways. In this paper, we describe our efforts at building a graphical database of neural connection patterns that preserves as much anatomical detail as possible.

Among current methods for representing neural connectivity information in a unified or summarized format, one of the most popular is the schematic wiring diagram in which boxes that represent different brain areas are connected with lines that denote the existence of connections between them (e.g. Felleman & Van Essen 1991). Although such diagrams are helpful in ascertaining global trends in connectivity, they gloss over a tremendous amount of information that is available in the data on a finer scale. For example, in some instances a fairly localized topographical organization to the connection patterns may exist (such as between areas V1 and V2) and in other cases not (such as between V4 and the inferotemporal complex). In addition, there may also be differences in the degree of connectivity. Some regions may be densely interconnected (showing heavy labelling), whereas others may be sparsely interconnected (showing only light labelling). This type of information can be captured to some extent by a connectivity matrix, in which each element denotes the strength of an inferred connection (e.g. Young 1993; Stephan *et al*. 2000*b*) but still only at the macroscopic level. Schematic wiring diagrams also typically do not differentiate between evidence against a connection versus lack of evidence—i.e. if a line does not connect two boxes in the diagram, it is not

*Author for correspondence (baolshausen@ucdavis.edu).

immediately obvious whether it is due to there being no labelling observed in an experiment that would reveal such a connection or whether the experiment simply has not yet been done.

For the investigator braving a foray into the literature in order to learn the specifics of certain connection patterns, there is yet another problem lurking, namely, different authors often use different nomenclatures to describe the same brain region (e.g. Brodmann 1909; Felleman & Van Essen 1991). The pulvinar nucleus of the thalamus is an excellent example of where there is wide disagreement among authors as to the placement of borders delineating various subnuclei. What constitutes the lateral pulvinar to one author may be the medial pulvinar to another. Other regions, such as inferior pulvinar, are constantly undergoing further subdivision (Cusick *et al.* 1993; Gary *et al.* 1999). Thus, one cannot simply trust a verbal report in a paper claiming to have found 'evidence for connections between V1 and lateral pulvinar'. One must examine the scored cross-sections and compare the actual label sites with one's own 'mental database' of these various regions. Although such work is routine for a skilled neuroanatomist, it cannot reasonably be expected of the wider class of investigators that need access to detailed information about neural connectivity. Thus, there is a need for tools that allow for the objective analysis and comparison of fine-scale neuroanatomical data.

Our effort to build a neuroanatomical database was initially stimulated by an interest in the patterns of connectivity between the pulvinar and visual cortex in the macaque monkey. The difficulties encountered in integrating information across multiple studies, due to the multiplicity of partitioning schemes used for both cortex and the pulvinar, motivated our development of a graphically orientated database. The idea behind a graphical database, as opposed to a conventional text orientated database, is that data are represented within a canonical neuroanatomical atlas, irrespective of partitioning schemes. Our database program, called XANAT, allows injection and label sites from multiple experiments and across different laboratories to be entered onto a single, consistent graphical representation of the anatomical structures of interest. The connectivity trends within the entire set of data may then be revealed using analysis routines, which display the cumulative evidence for connectivity between user-specified regions of interest in a colour-coded 'heatmap' that is superimposed on the neuroanatomical atlas. In this way, one can summarize the evidence for connectivity between any arbitrarily chosen locations in the brain, accumulated from all data within the database, without relying upon (though possibly guided by) previous area-partitioning schemes.

Graphical representations of data have found success in a number of anatomical applications, including volumetric atlases (Toga 1989; Jones 2000) and volumetric reconstruction of data collected from a series of slices (Schwaber *et al.* 1991; Funka-Lea & Schwaber 1994). In addition, graphical representations at the macroscopic level have been shown to allow for the translation of one parcelization scheme to another (Stephan *et al.* 2000*a*). Perhaps most closely paralleling our own efforts is the program NeuARt (Dashti *et al.* 1997), which allows sites of various types of anatomical labelling (not necessarily connectivity related) to be scored within a canonical atlas. The implementation of graphical search capability for this program is still in progress.

Another capability that we have attempted to build into our database is the ability to represent and combine data probabilistically. This capability is important, because there are uncertainties inherent in the data due to both the probabilistic nature of dye transport and the process of registering injection and labelling sites within the canonical atlas. Thus, rather than simply superimposing multiple datasets, one would actually like to obtain a measure of the probability that a connection pathway exists between one area and another, given all the data available. XANAT does this by using the rules of Bayesian inference to combine evidence from multiple datasets, thus taking into account both the structure and variability inherent in the data in a principled manner (Grenander & Miller 1993). Such a probabilistic approach has recently been employed successfully for combining information about areal boundaries in the cortex from different animals (Van Essen *et al.* 2001).

In this paper, we describe the structure of XANAT and the tools we have built for analysing data on neuroanatomical connectivity (see § 2). We then demonstrate its use in inferring patterns of connectivity between the cortex and pulvinar of the macaque monkey (see § 3). Finally, we discuss the lessons learned from this effort and some future directions for improving the methods of data entry and data representation.

2. METHODS

(a) *The graphical representation*

Injection and tracer data are represented in XANAT by their coordinates within a canonical brain atlas. This representation comprises three separate levels: the atlas, corresponding supplementary images, and explicit area-partitioning schemes.

The atlas comprises a set of images onto which all data are drawn, providing a canonical representation for combining results across experiments and studies. Ideally, these images are like those of a conventional atlas: they show the structural geography and cytoarchitectural organization and do not rely upon explicit segmentation. The atlas used in our exemplar of corticopulvinar connectivity includes two types of images (figure 1*a*,*b*): a flat map of *Macaca mulatta* cerebral cortex geography (gyri and sulci), and eight coronal Nissl-stained sections through the *Macaca fuscata* pulvinar and associated subcortical regions (posterior to anterior displayed from left to right, by row). The *Macaca fuscata* pulvinar atlas was used instead of one for *Macaca mulatta* because it was the highest quality atlas available at the time of development. The cortical flat map was scanned directly from Felleman & Van Essen (1991) using a Microtek flat-bed scanner. The pulvinar images were digitized directly from Kusama & Mabuchi (1970) using a 512×512 pixel charge-coupled device (CCD) camera and thus include Kusama & Mabuchi's designated thalamic nucleus boundaries and labels. These images also included stereotaxic coordinates, which were then translated into pixels in the atlas.

The supplementary images used in XANAT are in register with the atlas images and can be toggled to be displayed in their stead. These supplementary images contain additional information that aids data entry and analysis. For example, our flat map of cortical geography (figure 1*b*) has a corresponding

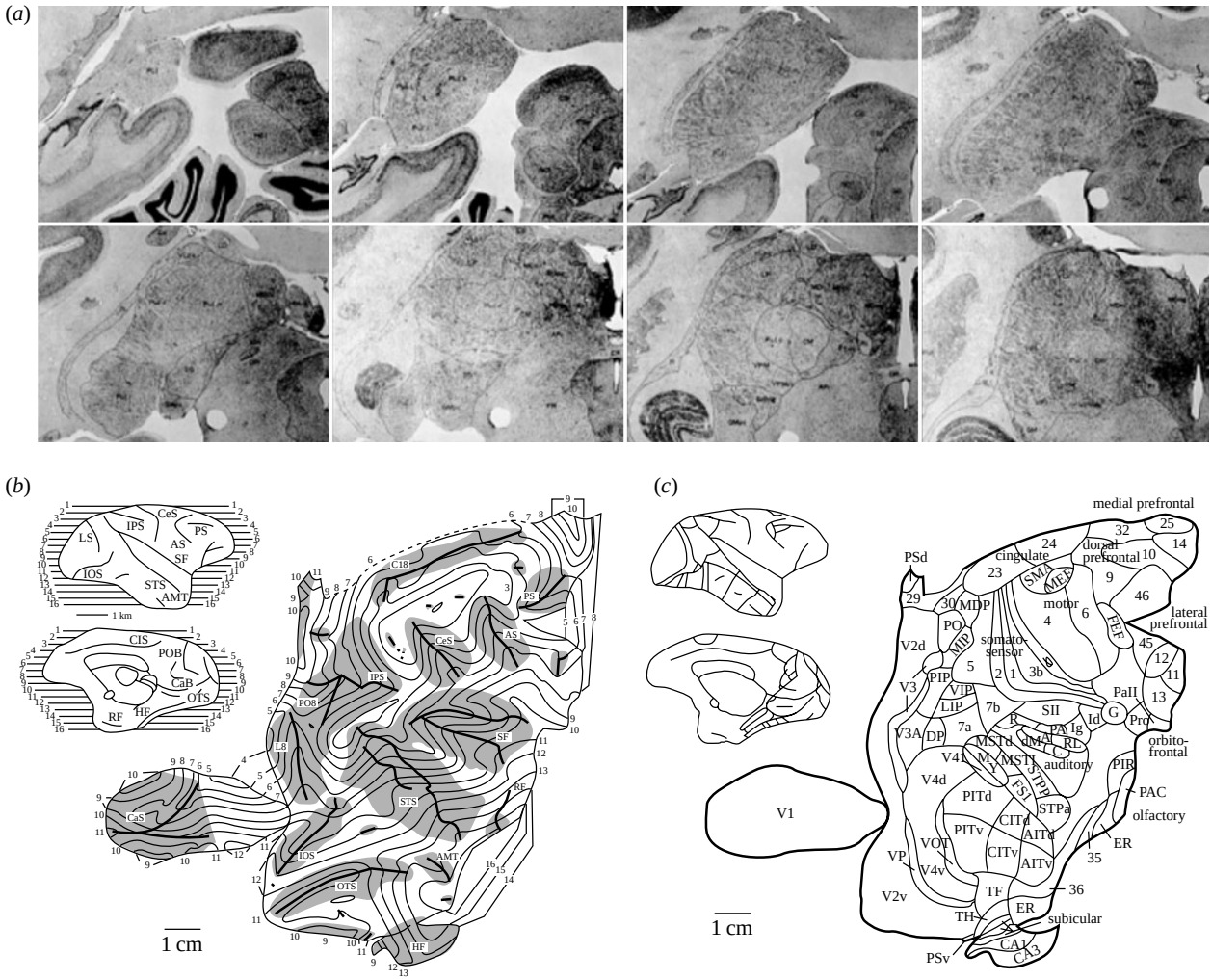

Figure 1. The atlas comprises images of the anatomical structures of interest. In this example, it consists of eight coronal Nissl-stained sections of the pulvinar and a flat map of macaque cortical geography. (*a*) The most caudal pulvinar section is shown on the top left, with increasingly rostral slices proceeding left to right and then down. (*b*) Atlas image of cortical geography taken from Felleman & Van Essen (1991). (*c*) Supplementary image for the cortical flat map showing area-partitioning scheme.

supplementary image of the Felleman & Van Essen (1991) area-partitioning scheme (figure 1*c*). We did not include supplementary images for the pulvinar sections because our atlas images already happened to include nucleus boundaries. Supplementary images for the pulvinar sections can easily be added, though, if one wished to use an alternate partitioning scheme, such as Cusick *et al.* (1993); Gary *et al.* (1999). Other possible supplementary images include myelin- or antibody-stained slices that correspond to Nissl-stained slices in the atlas.

To facilitate the interpretation of analyses performed in XANAT, it is useful to have an explicit segmentation of the atlas into its constituent subdivisions, not just an image of the partitioning scheme. This segmentation is achieved by drawing borders explicitly, as polygons, onto either the atlas or the corresponding supplementary images using an accompanying application tool. Although multiple partitioning schemes may be represented simultaneously, we limit our corticopulvinar example to just Felleman & Van Essen's (1991) partition scheme (corresponding to the supplementary image, figure 1*c*). These area-defining polygons are not displayed during normal XANAT use. However, overlap between these polygons and entered data are reported with every injection and label site drawn into the atlas, and can be used to guide data searches and interpret the results of analyses.

Images are typically limited to representing one hemisphere; however, the entire graphical representation (the atlas, supplementary images and area segmentations) can be flipped about their vertical axis. This facilitates the entry of data collected from either hemisphere.

(b) *Entering data*

Each data record in XANAT contains both text and graphical information that correspond to a single neuroanatomical experiment, showing a single injection site, multiple injection sites or a lesion, as well as the resulting pattern of labelling (or degeneration in a lesion study). These data, drawn as polygons and ellipses, are transferred manually from either illustrations or sectioned slices onto the atlas and/or supplementary images. Label strength and injection halos are encoded by stipple density and, in a multiple-injection study, different injections and labelling are distinguished from one another by colour (up to five different injections are supported in the current version of XANAT).

In addition to these graphical data, every record contains text information. These include the reference, tracer type, injection and labelling distribution by area (if areas are defined, see § 2a), comments and confidence in the data. The confidence allows the investigator that enters the data to evaluate subjectively its accuracy and assign a quantitative (0–100) measure.

This takes into account at least three factors: confidence in the original data, confidence in data entry, and confidence in tracer transport reliability. Because these assessments are subjective, it is not straightforward to develop a precise quantitative relationship for combining them into a single confidence value. Currently, we simply combine these factors mentally to come up with a single confidence score.

(c) *Analyses*

In addition to selecting each record separately for display and editing, XANAT contains analysis tools for combining data across injections and studies. There are two main types of analysis tools: list-generating analyses (searches and stacks), and graphical analyses that depict their results graphically (but also generate lists).

(i) *Searches and stacks*

The simplest analysis tools are searches and stacks, which reduce the original dataset by generating a subset list of records. The process by which these lists are created is dependent upon which method is used.

Searches select records that have certain keywords in the reference or comments field, as well as those that have projections to or from a particular area or areas. Search criteria can be combined using simple logical operators. For example, one can select a set of records that were published by a particular author, or a set that had connections with visual area V4 as well as some keyword of interest in the comments field, e.g. 'anterograde'.

Searching for records on the basis of area information requires an explicit segmentation of the atlas according to some area-partitioning scheme (discussed above). The search yields a list of records that show projections to or from the designated locations. The direction of projections in which one is interested determines whether injection or label is used to select a given record. For example, a search for inputs to area V4 will include those records that have either anterograde-tracer labelling or retrograde-tracer injections that lie in V4, because in either case the location of the injection or label, respectively, indicates the source of these projections. Conversely, records containing retrograde-tracer labelling or anterograde-tracer injections in V4 would not be included in the search results as they provide no information about projections to V4.

The stack list is set entirely by the user by selecting records individually. Items are pushed onto or popped from the stack using a LIFO (last in, first out) ordering. The stack primarily serves as a temporary store that facilitates the direct comparison of two or more otherwise unrelated entries. Also, the results of graphical analyses, described in the next section, can be pushed onto the stack, allowing one to compare various analyses with one another or with raw data.

(ii) *First-order graphical analyses*

Graphical analyses provide a way to extract and visualize information from multiple studies. The user specifies a region of interest within the atlas, and a graphical representation of connection strengths from that location, to that location, or both, is generated as a colour-coded 'heatmap'. Specifying the search area can be done either graphically or textually. Graphical search areas can be selected by drawing ellipses or polygons on any or all of the available images. One may also click on an existing polygon specified by an area-partitioning scheme. Areas having a corresponding text label can be selected by performing a text-based search and then using the results of the search for the analysis.

XANAT supports two primary methods of graphical analysis: superposition and probabilistic. Both yield a colour-coded heatmap, where colour represents a measure of the total strength (number of studies) or probability of connections, respectively.

A superposition analysis is generated using a weighted sum of the connectivity across the dataset. The relative weight each record contributes is determined by the product of three factors: the fraction of the search area that overlaps the record's relevant data (whether it be injection, or label, or both); the fraction of the record's relevant data that overlaps the search area; and the confidence weighting given the data at the time of data entry, or

$$\text{record weight} = \text{confidence} \times \frac{\text{size of (search} \cap \text{data)}^2}{\text{size of (search)} \times \text{size of (data)}}, \tag{2.1}$$

where size of (X) is the number of pixels in region X. The heatmap colour spectrum that results from this analysis ranges from least weight (blue) to greatest weight (red).

Although the superposition analysis offers a simple method for combining data across different studies, it suffers from the fact that it does not distinguish between a lack of evidence versus evidence against there being a connection. In addition, the heatmap value assigned does not properly reflect the total evidence for a connection because it will be heavily biased by the number of studies that happen to have been done for a particular region. The solution to these problems is to combine the data probabilistically.

In a probabilistic analysis, the resulting heatmap reflects the actual probability that each pixel in the atlas is connected with the identified search area, ranging from 0 (blue) to 1 (red), with a value of 0.5 indicating complete ignorance (green). Specifically, we calculate the posterior probability, $P(c(x,y)|D)$, where $c(x,y)$ is a binary-valued function that represents the presence or absence of a connection between the search area x and some point y in the atlas, and D is the set of relevant experimental data that speak to this connection, $D=\{d_1(x,y), d_2(x,y), \ldots, d_n(x,y)\}$. Each $d_i(x,y)$ is a binary variable that represents whether or not a connection is indicated by the ith study in the database (presence or absence of label). The posterior probability is given from Bayes' rule,

$$P(c(x,y)|D) \propto P(c(x,y))P(D|c(x,y)), \tag{2.2}$$

which states that the probability of there being a connection $c(x,y)$, given the data D, is proportional to the *a priori* probability of there being a connection $P(c(x,y))$ times the probability that the data accurately reflect the true state of connectivity, which for independent data is given by the factorial distribution

$$P(D|c(x,y)) = \prod_{i=1}^{n} P(d_i(x,y)|c(x,y)). \tag{2.3}$$

The values $P(d_i(x,y)|c(x,y))$ reflect the sensitivity and false-positive rate of the data, and in our particular implementation they are directly related to the user-assigned confidence in the data (see Appendix A).

Although the probabilistic approach does not obviate the need to make subjective judgements about the data, it does incorporate these subjective notions into quantitative reasoning in a consistent manner. We do not yet have a general solution for setting the sensitivity and false-positive rates $P(d_i(x,y)|c(x,y))$, and in fact the prescription we use here is rather crude and should be viewed merely as a starting point. Our main emphasis in this paper is on formulating a framework for combining these

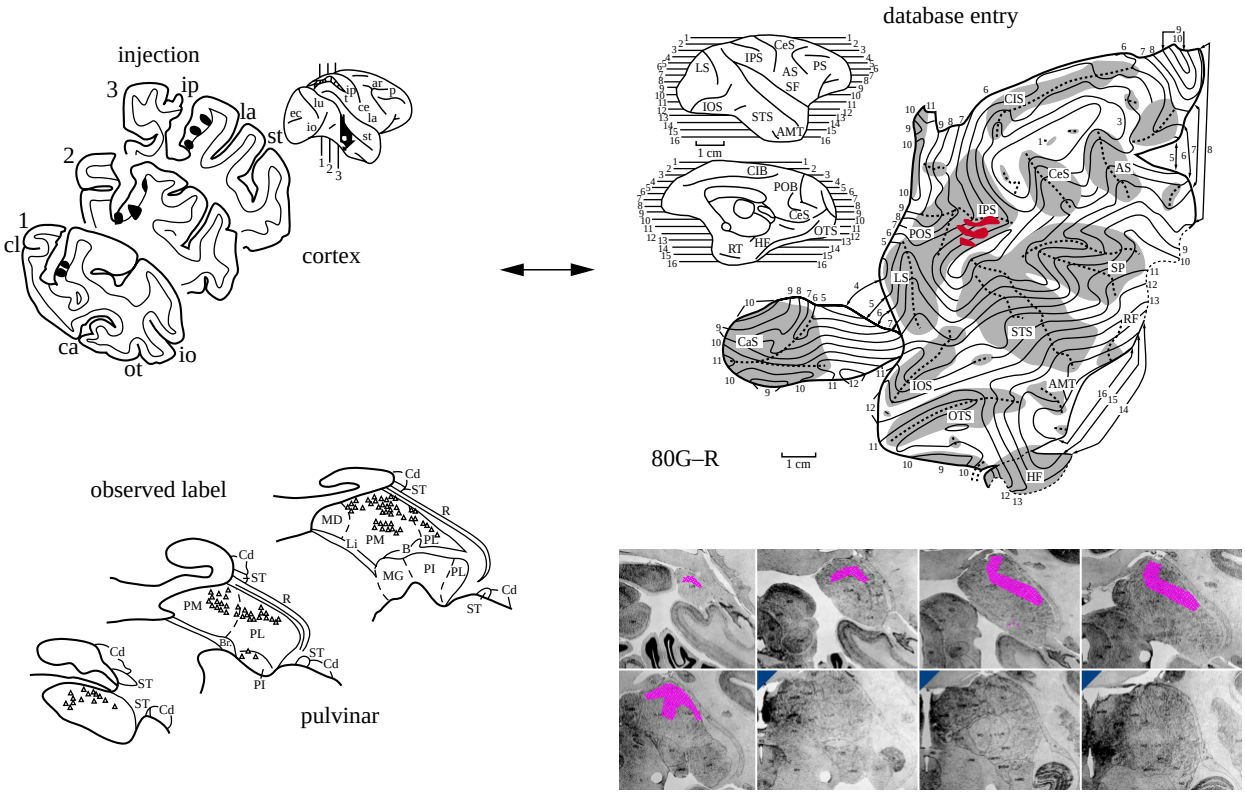

Figure 2. Data entry is performed by manually drawing sites of injection and observed labelling directly onto the neuroanatomical atlas. The entry of a cortical injection and resulting pulvinar label from a study by Baizer *et al*. (1993) is shown here. Atlas images not used in the original study are indicated by triangles in their upper-left corner.

uncertainties across multiple neuroanatomical studies, assuming they may be reasonably estimated.

(iii) *Second-order graphical analyses*

The results of the above first-order analyses may be accumulated, compared or contrasted by combining them together in various ways. These second-order analyses can use any graphical analyses as their substrate and come in three forms: additive, multiplicative and subtractive.

The additive analysis allows one to superimpose two analyses on top of one another, facilitating their comparison. The data from single records may also be added to a heatmap, allowing for direct comparisons between an analysis result and the data that generated it.

The multiplicative analysis provides a means of visualizing directly the amount of overlap between two analysis results. The geometric mean of the two operand analyses is taken on a pixel-by-pixel basis, and the colour-scale remains unchanged. In this way, analyses with little overlap will result in sparse non-zero values and multiplying an analysis with itself will leave it unchanged.

The subtractive second-order analysis illustrates the relative locations of two non-overlapping analyses. This is done by taking the pixel-by-pixel difference of two heatmaps and rescaling the colour-bar to accommodate the entire range, with zero in the middle.

3. RESULTS

(a) *Entering data*

Figure 2 shows the results of a retrograde-tracer injection in areas LIP (lateral intraparietal cortex) and VIP (ventral intraparietal cortex), from a study of Baizer *et al*. (1993), and its corresponding entry in the neuroanatomical atlas. The location of the cortical injection is represented by three separate polygons (red) within the cortical flat map. The zones where labelling was observed in the pulvinar, denoted by triangles in the figure, are represented by a series of polygons (purple) in each cross-sectional image, with some interpolation required in registering the cross-sections of the study with those in the atlas. To date, we have entered the results of 127 injection and lesion experiments in this way from 20 published studies. Our rough estimate of the accuracy with which tracer injection and labelling sites are typically represented is about 1–2 mm on average.

(b) *Analyses*

We used the database and the above-described analysis routines to investigate the overall trends in connectivity between visual cortex and the pulvinar. Figure 3*a* shows the results of analysing connections between cortical area V4 and the pulvinar using the superposition algorithm (the large image shows in more detail the fifth most caudal section of the pulvinar). This was done by selecting the V4 area-defining polygons (V4d and V4v) in the cortical atlas (or by typing 'V4' into the search window). The result of this analysis, accumulated over 21 relevant studies that contain injection or labelling within V4, reveals that most of these connections are with the inferior and lateral pulvinar nuclei at the dorsolateral zone of their shared border.

As mentioned in § 2, the superposition method provides a simple means of collapsing data across multiple experiments but it does not properly represent evidence for or

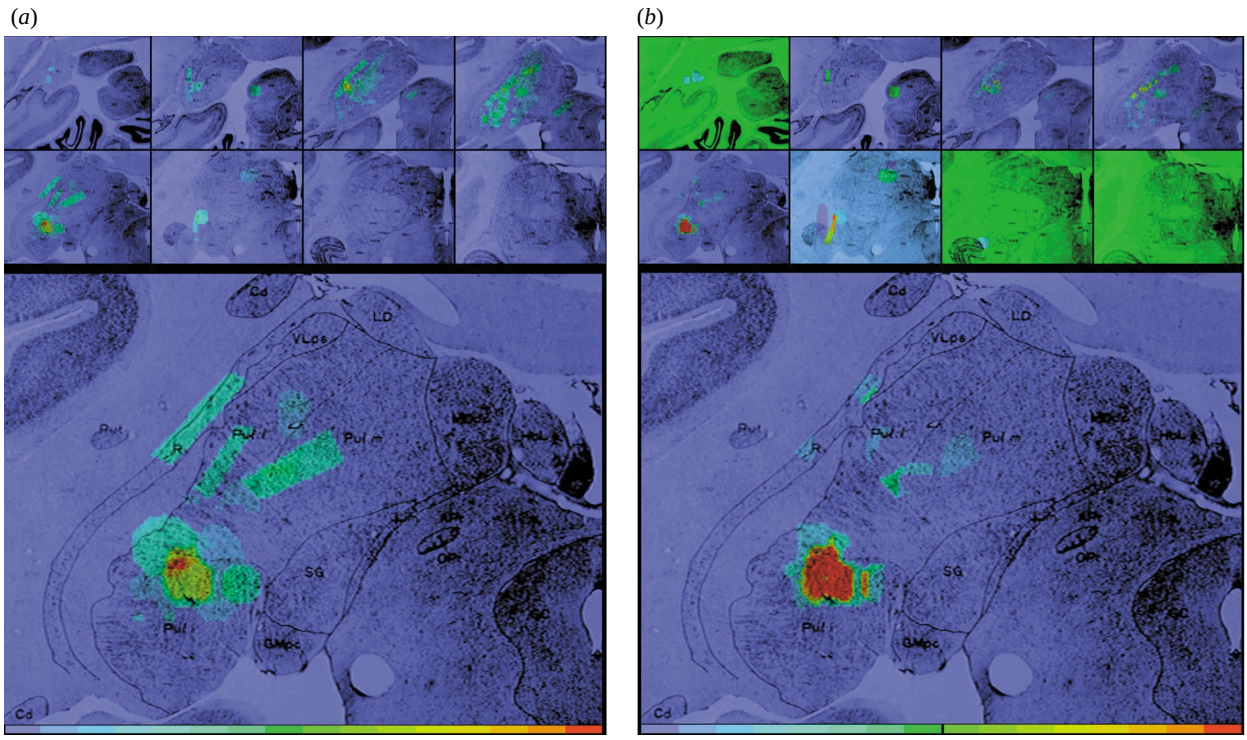

Figure 3. (*a*) Superposition analysis of connections between all of visual area V4 and the pulvinar. The left end of the colour spectrum (blue) represents minimal weight and the right end (red) represents the maximal weight. (*b*) A probabilistic analysis of connections between a small zone in the middle of V4d and the pulvinar. The left end of the colour spectrum (blue) represents a near 0% probability of there being such a connection and the right side (red) represents a near 100% probability; chance (green) is in the middle of the heatmap colour bar and is indicated by the black tick mark.

against connectivity in terms of probability. Thus, it is also useful to analyse the data probabilistically, as shown in figure 3*b*. Here, the heatmap represents the posterior probability of a connection with the search region, in this case a small subregion of V4. The green coloration visible in some of the pulvinar images now indicates that little information is actually available regarding visual cortical connections to the most rostral slices of the pulvinar (green corresponds to $P=0.5$). By contrast, the lack of connections between V4 and the more caudal medial pulvinar indicated by the superposition analysis appears to reflect an actual absence of connectivity. (To create this map, we assumed a value of 0.5 for the prior (see Appendix A), which assumes no prior knowledge of connections between the cortex and pulvinar beyond what is stored in the database.)

One of the principal uses of XANAT is to compare connectivity zones for different parts of the brain, such as different visual cortical areas. In figure 4*a*, we show the result of a superposition analysis for connections between the pulvinar and the ventral stream (V4/IT (inferotemporal cortex)). Within the pulvinar, these connections consistently appear in the inferior and ventral–lateral nuclei, indicating a role for these regions in form processing. This pattern of connections is in contrast to that found between the pulvinar and posterior parietal cortex (figure 4*b*). Here we find connections most predominantly within the medial pulvinar and more dorsal portions of the lateral pulvinar, indicating a role for these nuclei in processing spatial information (corresponding to the so-called 'where' stream). These patterns of connectivity illustrate how different pulvinar nuclei may be involved in processing different kinds of visual information (such as form versus spatio-temporal).

Different cortical areas within the same processing stream can also have different patterns of pulvinar connectivity. In figure 5, we use superposition analysis to examine how visual area V4 (figure 5*a*, reproduced from figure 3*a*) and inferotemporal cortex (figure 5*b*), both of the 'form' processing stream, are connected with the pulvinar. These small panels show that the connections between inferotemporal cortex and pulvinar are significantly more caudal and ventral than those between area V4 and pulvinar. The difference in locations of these two projections is highlighted by a subtractive second-order analysis (figure 5*c*), in which the results shown in figure 5*b* are subtracted from the results shown in figure 5*a*. Blue regions, such as ventromedial caudal pulvinar, indicate where inferotemporal cortex connects with the pulvinar more consistently than does area V4; red regions, such as more lateral and rostral pulvinar, indicate where V4 connects with the pulvinar more consistently than does inferotemporal cortex. Figure 5*d* shows a multiplicative second-order analysis, formed by multiplying the results shown in figure 5*a* with the results shown in figure 5*b*. Most regions in this multiplicative second-order analysis are near zero, indicating little overlap between these two sets of projections.

4. DISCUSSION

(a) *Analyses*

One method for analysing a collection of studies is to examine each study individually, developing a 'mental

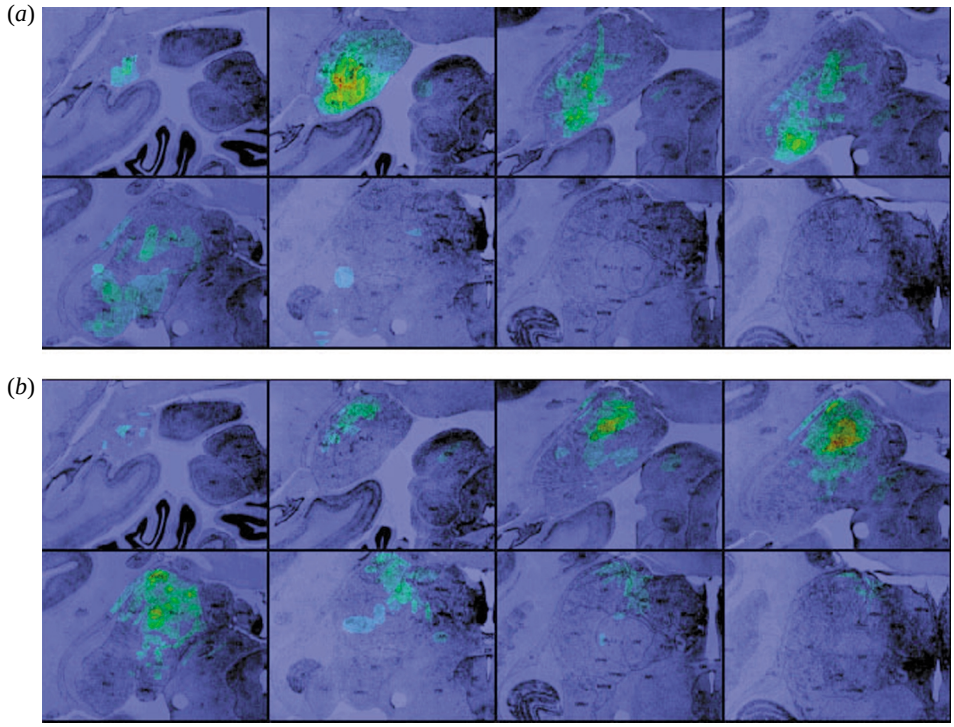

Figure 4. A first-order analysis comparison of pulvinar connections with ventral and dorsal visual processing streams. (*a*) Superposition analysis of ventral-stream (V4/IT (inferotemporal cortex)) connectivity with the pulvinar shows most of these connections are with the inferior and ventral–lateral nuclei of the pulvinar. (*b*) By contrast, posterior parietal cortex connects mostly with medial and dorsal–lateral nuclei. This comparison illustrates how different pulvinar nuclei may have a role in distinct processing streams.

model' of similarities and differences across the entire set. Such an analysis is certainly possible using XANAT; in fact, it is facilitated by all of the data being in a single, common representation. However, human memory is limited and subject to personal biases. XANAT can thus aid the process of understanding connectivity trends through analysis routines that combine data across many different studies in an objective fashion.

The superposition analysis is computationally simpler, and substantially faster, than the probabilistic analysis. This is because a superposition analysis examines each injection only once per analysis and the size of the search area has no effect on how long the analysis takes. By contrast, the probabilistic analysis evaluates separately the connectivity of every point in every image with every point in the search area; a search area larger than a few pixels thus requires an impractical amount of time to analyse. Furthermore, because of the nature of the computations involved, and because each pixel in the search area generates its own set of probabilities, combining these probabilities over a large number of search area pixels can potentially dilute the result. For this reason, probabilistic analyses should generally be constrained to small search areas (such as a single point or pixel in an area), whereas superposition analyses can be effectively used for analysing connections with entire areas.

The probabilistic analysis does offer a significant advantage over the superposition analysis in two main respects: (i) it allows one to distinguish evidence against a connection versus a lack of evidence for or against a connection; and (ii) it combines multiple datasets in a way that reflects properly the total evidence for connectivity, rather than the number of studies *per se*. This is not the case with a superposition analysis, since all heatmap values start at zero and data are combined additively. Thus, no data would appear the same as data showing evidence against a connection. In addition, if there happen to be many more studies for one particular area (e.g. more IT injections than V1 injections), then the results of the analysis will be significantly skewed in that direction. In the probabilistic analysis, by contrast, the heatmap value reflects the probability of there being a connection (0.5 being chance). Thus, the lack of observed labelling lowers the probability, whereas those regions for which no observations exist leave the probability unaffected. Also, although the observation of labelling increases the probability of a connection, multiple studies do not combine additively but rather asymptote towards 1.0. These distinctions are important because most studies do not include information for all regions of the atlas, and not all regions of the atlas are covered uniformly by the data.

Conversely, a probabilistic analysis requires that one has properly assigned uncertainties to the data, and so the overall conclusions resulting from such an analysis are only as good as the assumptions used in assigning the uncertainties in the first place. The particular method for assigning uncertainties we have chosen here is only intended as a beginning and could conceivably be improved upon if there were a principled basis for assigning them based on known uncertainties in the data (e.g. information about sensitivity or false-positive rates of dye transport from a particular study).

Second-order analyses provide the added flexibility of combining and comparing different analysis results, including past second-order analyses. The additive second-order analysis allows one to superimpose on analysis

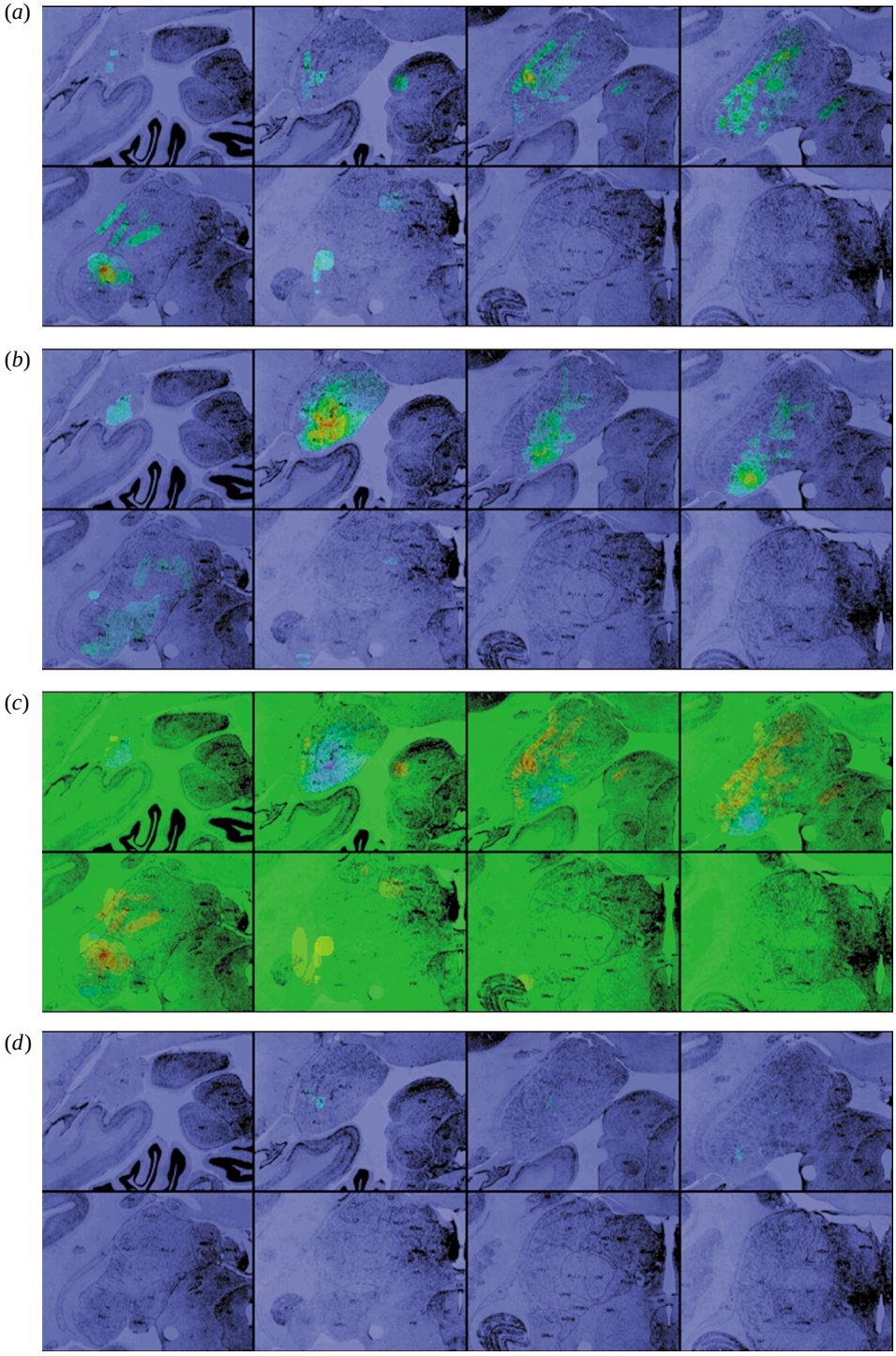

Figure 5. A first-order and second-order analysis of pulvinar connections with different levels of the ventral processing stream. (*a*) The superposition analysis of V4–pulvinar connections (same as figure 3*a*). (*b*) A superposition analysis of visual area IT (inferotemporal cortex) shows connections with more caudal and ventral zones of the pulvinar. (*c*) A subtractive second-order analysis highlights the difference between these two projections. (*d*) A multiplicative second-order analysis illustrates only a small amount of overlap.

results the original data that generated them; however, they do not provide as strong a contrast between two analyses. The subtractive second-order analysis shows more clearly the relative connections of two analyses. One caveat, though, is that a difference of zero could result from any two identical results, whether or not they indicate evidence for connectivity. These two alternatives can be distinguished using the multiplicative second-order analysis, which displays exclusively those areas with connections coincident between the two original analyses. Given the different capabilities of each second-order analysis technique, they are often best used in concert with one another to illustrate a given comparison.

Using these first-order and second-order analysis tools, distinct subdivisions within a structure can be distinguished by connectivity, independent of anatomical features. Furthermore, they can be used to discover novel connections not immediately apparent in the original

data and to suggest possible areas of interest for future investigation.

(b) *The graphical representation*

The results of the above-described first-order and second-order analyses are meaningful only in so far as the database entries accurately reflect the original data. One advantage of the graphical representation used by XANAT is that the representation of neuroanatomical data is not dependent on how the structures involved are partitioned into areas. This is significant, as area designations often vary between laboratories and can have a large amount of variability. Also, relying on area designations to analyse data for connection-based area designations reveals a troubling circularity.

A textual alternative to area-based spreadsheet tabulation is to represent data by their corresponding coordinates. This technique is commonly used in functional imaging studies and databases (Fox *et al.* 1994), in which regions of activation are described by the Talairach coordinates of each region's centre. A limitation of this technique, though, is that identical coordinates in two individuals do not necessarily correspond to identical structures. This is because spatially defined coordinate systems do not account for individual variability.

Although a graphical representation overcomes some of these problems, it has its own set of limitations. One limitation of XANAT is that not all studies are equally amenable to being transferred onto a given atlas. In our pulvinar–cortex implementation, for example, the atlas consists of eight coronal slices of the pulvinar spaced by 0.5 mm. However, some data may come from slices spaced by different distances and sectioned at different angles (possibly even in different planes). This variability is partially accounted for by the confidence measure assigned to each record.

Another limitation is the resolution of the atlas. As the data come from a number of different studies and animals, and are entered into one canonical reference frame, it is likely that the fine structure in connectivity patterns will become blurred. For example, Hardy & Lynch (1992) show that projections from pulvinar to cortical areas LIP and 7a arise from intercalated horizontal zones in the medial pulvinar. If a number of such injections are entered into the database out of phase with one another, this fine lamination will probably not be seen in analysis results, though the location of LIP and 7a projections, as a group, will be.

(c) *Future development*

XANAT's function ranges from a simple anatomical database to an analytical tool capable of educing simple, intuitive results out of a complex dataset. Although it has proven to be useful for comparing connectivity between different structures of the brain, it should be viewed as an initial effort that has room for growth and improvement.

One important area for development is to represent volumetric structures, such as the pulvinar, volumetrically, instead of as a series of slices. In doing so, data presented as slices could be incorporated into the atlas volume at its appropriate angle; presently, we are limited to the atlas's planes of section. Furthermore, analyses could be performed volumetrically instead of being limited to slice surfaces. A volumetric representation would also be more informative and easily interpretable than the current slice representations and would provide the opportunity to view arbitrarily orientated planes of section.

An additional area for future development is automated data entry. Instead of manually mapping published or collected data onto our canonical representations, a warping algorithm (such as Christensen *et al.* 1997) could be used to automate the procedure. Recent anatomical software packages, such as reported in Funka-Lea & Schwaber (1994) and Van Essen *et al.* (1998), have started incorporating this technique. How such warping would be implemented in XANAT depends, in part, on whether the canonical structure is represented by slices (as in the current version) or by a volume. Regardless of the particular implementation, though, a warping algorithm would provide the means to transfer anatomical data into the database in an objective and consistent manner, and would also facilitate the sharing of these data with other databases and atlases. Furthermore, any distortions introduced by warping could be directly incorporated into our measure of confidence.

Finally, we have not dealt up to this point with the issue of quality control, as our emphasis has been to first address the issues of representation and inference of neural connectivity *per se*. Clearly, if XANAT is to be used as a universal repository of neural connectivity information for the greater scientific community, then some means must be taken to control submissions to the database, such as approval by referees (similar to the process of publishing a paper). In addition, standards should be adopted for assigning confidence values and uncertainty to data, for example by having a standard set of exemplar data available for which confidence values have been agreed upon by a panel of experts. Thus, future development will require providing the means to document more carefully the data entry and approval process in order for XANAT to serve as a general neuroscience resource.

5. CONCLUSIONS

XANAT provides an objective framework for combining data from different studies and from different laboratories. Data entry is independent from areal distinctions—distinctions that may be based on cytoarchitecture or myelination and not functional organization—and its analyses yield intuitive and easily interpretable results. These results, coalesced from a large pool of data, may provide insight into a structure's connectivity in ways that may not be obvious from an unaided search of the literature and that may lead to area distinctions based on functional organization. It is our hope that this database proves useful for researchers interested in the brain's anatomical connectivity and for database developers interested in creating the tools necessary for its study.

6. AVAILABILITY

XANAT is available at http://redwood.ucdavis.edu/bruno/xanat/xanat.html. It requires a UNIX workstation running the X11 window system, as well as a modifiable 8-bit colour map. Use of the program is free and unrestricted,

and independent development is encouraged. It is requested that acknowledgement be provided in proportion to what has already been written.

APPENDIX A: PROBABILISTIC INFERENCE OF CONNECTIVITY

Given some data, $d(x,y)$, which provides evidence for a connection between x and y, the posterior probability of there actually being a connection $c(x,y)$ is denoted $P(c(x,y)|d(x,y))$, where $P(A|B)$ means 'probability of A given B'. This is computed from Bayes' rule:

$$P(c(x,y)|d(x,y)) = \frac{P(c(x,y)))P(d(x,y)|c(x,y))}{P(d(x,y))}. \quad \text{(A1)}$$

The denominator, $P(d(x,y))$, serves here mainly as a normalization constant and may be ignored.

The prior $P(c(x,y))$ expresses the belief that a projection exists from x to y before the data are taken into account. If nothing is known beforehand, then $P(c(x,y))$ is set to 0.5, meaning that it is equally probable that a projection does or does not exist from x to y. If there is prior knowledge that alters this prior belief, it can be incorporated by raising or lowering $P(c(x,y))$ appropriately.

The likelihood function $P(d(x,y)|(c(x,y))$ represents the probability that the data $d(x,y)$ would have arisen from a given connectivity state $c(x,y)$. A complete lack of confidence in the data would yield a likelihood function value, $P(d(x,y)|(c(x,y)=\mu)=0.5$, meaning that the data are completely uninformative about the true connectivity $c(x,y)=\mu$ ($\mu=0$ or 1). Values approaching 1 indicate increased assuredness that the data accurately reflect the true underlying connectivity, whereas values approaching 0 indicate increased belief that the data represent the exact opposite of the true state of affairs.

Note that the likelihood function is not necessarily symmetrical, i.e. $P(d(x,y)=1|c(x,y)=1)$ is not necessarily equal to $P(d(x,y)=0|c(x,y)=0)$. Letting τ represent successful label transport ($d(x,y)=1$) and χ represent an underlying connection ($c(x,y)=1$), then $P(\tau|\chi)$ conveys the probability that label is transported given that there is a connection present (sensitivity), whereas $P(\overline{\tau}|\overline{\chi})$ conveys the probability that label is not accidentally transported to a location, given that there is no connection present ($P(\tau|\overline{\chi})=1-P(\overline{\tau}|\overline{\chi})$ is the false-positive rate). These two probabilities describe different physical processes, e.g. the probabilistic nature of dye transport, and therefore in general warrant two separate confidence measures. $P(d(x,y)|c(x,y))$ is thus given by the following table:

		$c(x,y)$	
		0	1
$d(x,y)$	0	$P(\overline{\tau}\|\overline{\chi})$	$1-P(\tau\|\chi)$
	1	$1-P(\overline{\tau}\|\overline{\chi})$	$P(\tau\|\chi)$

In our particular implementation, we take $P(\overline{\tau}|\overline{\chi})=P(\tau|\chi)$, which is equivalent to assuming that the false-positive and false-negative rates are identical. (Given that one of the major sources of uncertainty in our data is in properly aligning data within the atlas, this is fairly appropriate.) Thus, we determine these values directly from the confidence values assigned to the data according to the following formula:

$$P(i|i) = \frac{\frac{\text{confidence}_{i|i}}{2}+50}{100}. \quad \text{(A2)}$$

To combine these uncertainties across multiple studies, $d_1 \ldots d_n$, each piece of data is assumed to be independent, which is a reasonable assumption for the type of data we are dealing with, and so

$$P(d_1(x,y), \ldots, d_n(x,y)|c(x,y)) = \prod_{i=1}^{n} P(d_i(x,y)|c(x,y)). \quad \text{(A3)}$$

Additionally, each piece of data only provides information for a limited region of x- and y-values, and so one's belief in $c(x,y)$ should only be updated for those particular points. If U denotes the region of injection uptake and L denotes the valid region in which to look for label (e.g. outside the halo of injection), then, for an anterograde tracer, $P(d(x,y)|c(x,y))$ is defined for all $x \in U$ and $y \in L$, whereas for a retrograde tracer datum it is defined for all $x \in L$ and $y \in U$.

Finally, to calculate the total probability for a given search region, we calculate the average probability of any point within the search area being connected to any other point in the atlas:

$$\begin{aligned} P(\text{search} \rightarrow y) &= \frac{1}{\text{area(search)}} \sum_{x \in \text{search}} P(c(x,y)|d(x,y)) \\ P(x \rightarrow \text{search}) &= \frac{1}{\text{area(search)}} \sum_{y \in \text{search}} P(c(x,y)|d(x,y)) \end{aligned}, \quad \text{(A4)}$$

where $a \rightarrow b$ denotes 'a connects to b'.

We thank Chris Lee and Harold Burton for helpful advice during the building of the database. Supported by ONR AASERT grant no. 11 SP/93/R0334 (W.A.P.) and National Institutes of Health grant no. MH57921 (B.A.O.).

REFERENCES

Baizer, J. S., Desimone, R. & Ungerleider, L. G. 1993 Comparison of subcortical connections of inferior temporal and posterior parietal cortex in monkeys. *Vis. Neurosci.* **10**, 59–72.

Brodmann, K. 1909/1994 *Localisation in the cerebral cortex*. London: Smith Gordon. [Translated by L. J. Garey.]

Christensen, G. E., Joshi, S. C. & Miller, M. I. 1997 Volumetric transformation of brain anatomy. *IEEE Trans. Med. Imag.* **16**, 864–877.

Cusick, G. C., Scripter, J. L., Darensbourg, J. G. & Weber, J. T. 1993 Chemoarchitectonic subdivisions of the visual pulvinar in monkeys and their connectional relations with the middle temporal and rostral dorsolateral visual areas, MT and DLr. *J. Comp. Neurol.* **336**, 1–30.

Dashti, A. E., Ghandeharizadeh S., Stone J., Swanson L. W. & Thompson, R. H. 1997 Database challenges and solutions in neuroscientific applications. *NeuroImage* **5**, 97–115.

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

Fox, P. T., Mikiten, S., Davis, G. & Lancaster, J. L. 1994 BrainMap: a database of human brain mapping. In *Functional neuroimaging* (ed. R. W. Thomas), pp. 95–105. San Diego: Academic Press.

Funka-Lea, G. D. & Schwaber, J. S. 1994 A digital brain atlas and its application to the visceral neuraxis. *J. Neurosci. Meth.* **54**, 253–260.

Gary, D., Gutierrez, C. & Cusick, C. G. 1999 Neurochemical organization of inferior pulvinar complex in squirrel monkeys and macaques revealed by acetylcholinesterase histochemistry, calbindin and Cat-301 immunostaining, and *Wisteria floribunda* agglutinin binding. *J. Comp. Neurol.* **409**, 452–468.

Grenander, U. & Miller, M. I. 1993 Representations of knowledge in complex systems. *J. R. Statist. Soc.* B **56**, 549–603.

Hardy, S. G. & Lynch, J. C. 1992 The spatial distribution of the pulvinar neurons that project to two subregions of the inferior parietal lobule in the macaque. *Cereb. Cortex* **2**, 217–230.

Jones, E. G. 2000 *Human brain project brain atlas*. See http://neuroscience.ucdavis.edu/HBP/.

Kusama, T. & Mabuchi, M. 1970 *Stereotaxic atlas of the brain of* Macaca fuscata. University of Tokyo Press.

Schwaber, J. S., Due, B. R., Rogers, W. T., Junard, E. O., Sharrna, A. & Hefti, F. 1991 Use of a digital brain atlas to compare the distribution of NGF- and bNGF-protected cholinergic neurons. *J. Comp. Neurol.* **309**, 27–39.

Stephan, K. E., Zilles, K. & Kötter, R. 2000*a* Coordinate-independent mapping of structural and functional 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., O'Neill, M. A., Young, M. P. & Kötter, R. 2000*b* Computational analysis of functional connectivity between areas of primate cerebral cortex. *Phil. Trans. R. Soc. Lond.* B **355**, 111–126.

Toga, A. W. 1989 Digital rat brain: a computerized atlas. *Brain Res. Bull.* **22**, 323–333.

Van Essen, D. C., Drury, H. A., Joshi, S. & Miller, M. I. 1998 Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proc. Natl Acad. Sci. USA* **95**, 788–795.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B. H., Bakircioglu, M. & Miller, M. I. 2001 Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* **41**, 1359–1378.

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