Knowledge management of the neuroscientific literature: the data model and underlying strategy of the NeuroScholar system

Gully A. P. C. Burns

Department of Neurobiology, University of Southern California, Los Angeles, CA 90089-2520, USA (gully@usc.edu)

This paper describes the underlying strategy and system's design of a knowledge management system for the neuroscientific literature called 'NeuroScholar'. The problem that the system is designed to address is to delineate fully the neural circuitry involved in a specific behaviour. The use of this system provides experimental neuroscientists with a new method of building computational models ('knowledge models') of the contents of the published literature. These models may provide input for analysis (conceptual or computational), or be used as constraint sets for conventional neural modelling work. The underlying problems inherent in this approach, the general framework for the proposed solution, the practical issues concerning usage of the system and a detailed, technical account of the system are described. The author uses a widely used software specification language (the Universal Modelling Language) to describe the design of the system and present examples from published work concerned with classical eyeblink conditioning in the rabbit.

Keywords: neuroinformatics; data model; knowledge management system; published literature

1. INTRODUCTION

Neuroscientists suffer from information overload due to the extent, the complexity and the complicated taxonomy of their subject. This situation was explicitly recognized in the first sentence to appear in the journal *Trends in Neuroscience*: 'Even the most active neuroscientist spends more working hours in reading, reviewing and writing scientific reports than on direct experimental effort' (Bloom 1978). Seventeen years later, the same author reaffirmed his previous claim and anticipated the development of 'knowledge management systems' in order to alleviate this problem (Bloom 1995).

Within the field of informatics, the expression 'knowledge management' carries a specific meaning. If 'data' are considered as isolated, simple facts, and 'information' is data that have been sorted, analysed and interpreted, then 'knowledge' is information that has been placed in the context of other known information (Blum 1986). Within the context of the published literature, these concepts may be organized in the following way. Each individual observation in the results and methods sections of a given article comprises a single datum. The integration of all data into a single organized data model forms the paper's information. Finally, the interpretations that authors and readers make of this information (when considered in the broader context of the literature as a whole) form the paper's knowledge.

The global scope of the complexity of knowledge in neuroscience is daunting; for example, neuroanatomical descriptions name over 1000 named structures (i.e. nuclei, cortical areas and undifferentiated regions of the brain) in the brain of the rat (Swanson 1998). Each one of these

structures may contain as many as 10 different cell groups (made up of thousands of individual cells) and the connections of these cell groups tend to be highly parallel across systems (Swanson 1995). Given that inclusion of dynamic or functional properties into our theoretical models of these systems will certainly add orders of magnitude to the complexity of the problem of understanding the whole system, there is a strong scientific demand for large-scale methods of information management.

Within this paper, the conceptual and technical design of the NeuroScholar system—a computational knowledge management system with the objective of providing non-computational neuroscientists with a method to manipulate the published literature—is described. In order to restrict the focus of the current project, NeuroScholar's current design is concerned with addressing the following scientific problem: to fully delineate the neural circuitry involved in a specific behaviour.

It is proposed that users may interact with the literature in a new way through this system, in order to generate and manipulate models of their knowledge. This is not altogether unlike other methods of model building within neuroscience (for example, see Goddard *et al.* (this issue)), but this approach is geared towards neuroscientists who do not explicitly use computational methods to theorize about their subject. These scientists tend to use a conceptual or argument-based approach that draws on the wealth of information within the literature (as well as from their own experimental work), but may find that the volume of information is too large to be able to study any system involving more than a handful of separate neural components.

When considering the number of man-hours taken in its generation, its financial cost and the significance of its contents, the value of the literature is inestimable, but, as a resource for researchers, its value is limited by its usefulness. Conceptual and logistical problems of collating and structuring this information become non-trivial when, as is the case with systems-level neuroscience, theoretical models must be constructed with information from a large number of different studies, under different experimental paradigms, with little or no standardization. Over the last decade, the use of computational database systems to manage and store publicly available data has involved two distinguishable approaches: local user-specific commercial tools and centralized systems.

The first serious attempt to generate a neuroinformatics database based on published (or 'public') neuroscientific data was probably the BRAIN BROWSER (Bloom *et al.* 1989). This multi-level, cross-domain, atlas-based encyclopaedia could be annotated with data, graphical icons and explanations in a structured way that made sense to neurobiologists. It was originally distributed as a Macintosh HyperCard stack acting as a non-centralized desktop application. Thus, data sharing was possible by exchanging (or exporting) files between installations of the software, but the application itself acted as a local repository of structured information based on the design of BRAIN BROWSER's data model. This program is a very general commercial tool that neuroscientists may use to structure their own interpretation of public data.

Researchers studying the system-level organization of neural circuitry adopted a different approach by collating large amounts of neuroanatomical connection data from the literature and publishing them as data summaries (with accompanying references to each cited publication) in order to obtain the input data for subsequent analysis (Nicolelis *et al.* 1990; Felleman & Van Essen 1991; Young 1993; Scannell *et al.* 1995, 1999; Burns 1997; Burns & Young 2000; Stephan *et al.*, this issue). For example, a hierarchical organization for the macaque monkey visual system was postulated by classifying individual connections as 'ascending', 'descending' or 'lateral' according to where the connection originated or terminated within the laminated structure of the cortex (Felleman & Van Essen 1991). Because the authors of this paper tabulated and presented all the data supporting their scheme, other workers were then able to re-examine the initial dataset using more sophisticated computational methods, finding over 200 000 hierarchies that fitted the anatomical constraints more accurately than the solution that had been calculated by hand (Hilgetag *et al.* 1996).

Other studies of the organization of neural systems in the rat, cat and monkey treated the neuroanatomical data as a graph (i.e. a set of vertices that are interconnected by edges) in order to use mathematical analysis (Nicolelis *et al.* 1990; Felleman & Van Essen 1991; Young 1992, 1993; Scannell *et al.* 1995, 1999; Burns 1997, 1999). Analyses of the connections of cat cortex enabled researchers to make predictions concerning physiological properties of the circuit that were then tested experimentally: the presence of plaid-pattern sensitive cells in the anterior ectosylvian sulcus of the cat was predicted in Scannell *et al.* (1995) and then confirmed experimentally in Scannell *et al.* (1996). The CoCoMac project is, at

present, the most sophisticated operational system involved in the dedicated collation of neuroanatomical and neurophysiological literature-based data and is concerned with investigating connectivity data in the macaque monkey (Stephan *et al.* 2000a, this issue).

NeuroScholar amalgamates these two approaches by providing the computational architecture for neuroscientists to define, build, share and analyse their own computational representations of published neuroscientific knowledge in a centralized system.

Previous development work in this field has highlighted several areas of risk for literature-based informatics solutions. There are four areas of major concern: the process of data entry is slow, laborious and highly subjective; the nomenclature tends to be highly non-standardized and contradictory; arguments tend to be constructed using complex language that are difficult to represent in a machine-readable format; and the literature comprises studies of variable importance and validity. These issues may be addressed at a design level (as is the case in this article), but really cannot be thoroughly tested or addressed in the absence of a fully working system.

Allowing multiple users to enter and manipulate data simultaneously may alleviate the time constraints of the process of data entry, whilst requiring that data entered by different users should be continuously tracked to keep track of potential subjective differences. Solutions to anticipated difficulties concerning the complexity of the data and nomenclature use a rule-based approach involving existing computational methods, such as the objective relational transformation (see below; Stephan *et al.* 2000b). It may be possible to alleviate this problem by automating parts of the process of data entry, either by using computational methods of data extraction from prose (Hobbs 1997), or by incorporating the process of data collation directly into the automated aspects of the publication process. At present, these solutions have not been implemented and the process of data entry remains the rate-determining step for the usage of literature-based databases.

The general problem of heterogeneity of terminology is illustrated by considering the neuroanatomical nomenclature, where standardization has been notoriously difficult to accomplish, despite attempts over the last century to do so (compare the work of Wilder (1896), from the 19th century, to recent studies by Bowden & Martin (1995)). A strong case may be made for why such attempts at standardization are ill-advised within neuroanatomy, specifically because the subject is still developing rapidly and most delineations of brain tissue are made with less-than-complete information (see Appendix VI, 'Problems with Nomenclature' in Swanson 1998). It is suggested that rather than defining a set of common high-level conceptual terms that neuroscientists are expected to adopt unequivocally, it may be more constructive to standardize the representations of the logical relationships that link regions of brain tissue or designated neuronal populations. Stephan and Kötter developed the objective relational transformation (ORT) as a software-based methodology to accomplish this for monkey cortical parcellation schemes by comparing schemes in a highly constrained manner and building a logical framework of inference based on the spatial relationships between

cortical areas (Stephan *et al.* 2000b). This concept of defining sets of relations to link data entities across nomenclatures has been incorporated into NeuroScholar's design (see § 2a(ii)).

The functionality of the BRAIN BROWSER software application stems from the inclusion of the representation of a brain atlas within its user interface. This provides a graphical interface that can resolve many issues straightforwardly involving neuroanatomical parcellation schemes as well as being a more intuitive framework for users than a purely text-based system (as is the case with all other systems mentioned above). The use of atlases within experimental neuroscience provides an essential global neuroanatomical framework where data from different experiments may be brought into register. The importance of these publications may be demonstrated by considering the citation index of widely used brain atlases. For example, the number of citations of the first and second editions of Paxinos and Watson's atlas of the rat brain (Paxinos & Watson 1982, 1986) had a combined citation index in excess of 17 000 within the Institute for Scientific Information's online citation database in December 2000. This should be compared with the citation index of roughly 4000 for Hubel and Wiesel's Nobel award winning publication in 1962 (Hubel & Wiesel 1962). Thus, an important aspect of the functionality of a literature-based, knowledge base management system involving neuroanatomy is for users to be able to embed their interpretations into the spatial framework of widely used atlases without restricting them to a single specific reference scheme. Within the current design phase of NeuroScholar, Larry Swanson's second edition atlas of the rat brain is used (Swanson 1998) but the concepts may be generalized for use with any other atlas.

Neuroinformatics systems have been developed that allow users to collate experimental data into atlases (for one example, see Press *et al.* (this issue)). One such system was the NeuARt system. This is software used to analyse neuroanatomical data developed at the University of Southern California, USA, within the laboratory of Larry Swanson under funding from the Human Brain Project (Dashti *et al.* 1997; Shahabi *et al.* 1999). For example, when considering the results of a *Phaseolus leucoagglutinin* (PHA-L) tract-tracing experiment (Gerfen & Sawchenko 1984), workers examine the experimental tissue microscopically and draw individual PHA-L-stained fibres by hand onto a digital version of the atlas (Swanson 1998). Even though each drawing produced under this method is subjective, based on the expertise of the researcher that produced it, if enough care is taken, the quality and accuracy of this representation is extremely high. Within the current design of NeuroScholar, the capability to represent various types of spatial entities into a version of the computational version of a commercially available atlas is provided (Swanson 1998). Spatial entities within this context are sets of points, lines or bezier curves, voxelated datasets, delineations of regions or the tree-like set of lines used to represent neuronal branching patterns (see § 2a(iii)).

The quality and readability of data within the literature varies enormously. In some cases, text and figures contradict each other, figures may be omitted or they may fail to show important details. This problem is

specifically concerned with how well the published information supports the definition of a constrained model. The CoCoMac system addresses this problem by defining enumerated sets of 'precision description codes' (PDCs) for each variable of interest in their data model (e.g. labelling density, the definition of parcellation maps, see Stephan *et al.* (this issue)). Within NeuroScholar, the concept has been generalized to attach an ordinal measure of precision to each individual classification of data fragments from the literature within the system (see descriptions of 'fragment precision' association classes, § 2a(ii)).

The precise way that scientists use natural language to communicate their ideas may influence the interpretation that readers will draw from it. One of the important features of preparatory work for this project was the recognition that interpreting data from a paper required that the reader formulate an argument whilst reading the paper. If the reader collated the end product of that argument into a database without also recording the basis for the argument and his or her own reasoning supporting it, then trying to understand the basis for the interpretation in the database months or years later would be impossible without re-reading the paper and reconstructing the argument (Burns 1997; Burns & Young 2000). In order to avoid this issue, the explicit claim that literature-based neuroinformatics solutions must represent (or directly access) the original text of the papers they describe is made.

The scientific goal of the project is to describe comprehensively the neural circuitry underlying a specific behaviour as completely as the literature will allow. This goal can be broken down into three high-level stages: to identify and describe a behavioural phenomenon; to identify and describe the neural populations involved in the behaviour; and, finally, to describe all axonal pathways that exist between all previously identified neuronal groups.

The source of the information to support this process originates in publications describing different classes of experiment within the neuroscientific literature. The subject of ethology is based on a formal treatment of animals' behaviour based on hierarchical schemes to link high-level aspects of behaviour to individual movements, whole-body actions, or the physiological activity of individual effector organs and muscles (Weiss 1941). An important theoretical construction in ethology is the 'ethogram', which may be thought of as an account of the behavioural repertoire of an animal (Tinbergen 1951). Ethograms may be formulated as mathematical graphs consisting of a network of interconnected nodes that represents different states of activity of the animal and the possible transitions between those states (see Grant (1963) for an example of an ethogram for aggressive behaviour in the rat). This concept of graph-based ethograms forms the basis of NeuroScholar's treatment of behaviour.

The second stage of the high-level strategy involves determining which brain structures constitute the neural circuitry responsible for producing the behaviour under study. Experiments that are relevant to this issue may be classified into two categories. Experiments in 'observational physiology' involve the measurement of changes in physiological characteristics of specified neuronal populations

when the animal performs a behavioural paradigm. Experiments in 'interventional physiology' involve making a physiological intervention in specified neuronal populations to produce a measurable change in the animal's behaviour under the specified paradigm.

Observational physiological experiments that involve global mapping are particularly informative since physiological changes may be identified in several sites over large regions of brain tissue. Immediate early gene expression has been correlated to neural activity (Hunt *et al.* 1987) and topographical studies are often used to identify regions of putative activity under a given behaviour (Sheng & Greenberg 1990). These studies may use immunohistochemical staining to label expressed proteins (e.g. Canteras *et al.* 1997) or they may use *in situ* hybridization to label directly the mRNA involved in gene expression (e.g. Kollack-Walker *et al.* 1997). Other topographical approaches involve the detection of changes in brain metabolism (Morton *et al.* 1984), including modern scanning methods such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), even though these methods are rarely used to study non-human animals. Electrophysiological recording experiments are often used to examine temporal aspects of neuronal activity and this may be directly correlated to behaviour with chronically implanted microelectrodes that may be recorded from an awake, behaving animal (e.g. Nicolelis *et al.* 1997).

Studies of the behavioural effects of small discrete lesions are the most widely used physiological interventions used to investigate involvement of a region in neural processing underlying a specific behaviour. Radiofrequency, physical and thermal lesions damage neural fibres as well as neuronal cell bodies in the lesion site so that, even if an animal's behaviour is altered by the presence of a lesion, the somata of the neurons responsible for the changes may be in a region distant to the lesion site. So-called 'cytotoxic' lesions do not suffer from this problem by largely sparing fibres (Hastings *et al.* 1985). Cooling offers a reversible method of temporarily deactivating the neurons surrounding an implanted probe (Skinner & Lindsley 1968). Microcannulas or microelectrodes may be implanted chronically into a brain region in order to influence the activity of neurons in that region chemically or electrically in order to elicit changes in the animal's behaviour (Lammers *et al.* 1988).

The investigation of the structure of the neural circuitry running between different populations of neurons was revolutionized 30 years ago with the advent of tract-tracing experiments. This permitted researchers to identify the regions of origin and termination of long axonal connections (Lasek *et al.* 1968; Kristensson & Olsson 1971). These techniques have been significantly refined (see Kobbert *et al.* (2000) for a recent review) so that neuroanatomists can now describe connections in terms of the precisely defined positions of the somata and terminal fields of labelled cells on templates taken from a standard brain atlas (see Canteras & Swanson 1992; Risold *et al.* 1994).

The basic design of tract-tracing experiments is common across almost all methods. An experimental animal is anaesthetized and small amounts of tracer

chemicals are injected into the region of the brain under investigation. The tracer is taken up by the parts of neurons that extend into the injection site, typically at axonal terminals or dendritic arbours. The tracer is then transported along the cell's axon by intracellular transport mechanisms, either 'anterogradely' (from the cell body to the axonal terminals) or 'retrogradely' (in the opposite direction), or both. Finally, the animal is killed and its brain sectioned and stained to reveal a pattern that shows the distribution of transported label (Blackstad *et al.* 1981). The end product is a three-dimensional (3D) map of tracer concentration at the time of the animal's death. Since tracer is originally applied in a localized injection to a small region of the brain, labelled regions of the brain must be connected to the injection site. The data are usually presented as photographs or drawings of sections which show the distribution of label throughout the brain (Swanson 1998).

NeuroScholar addresses the task of collating and interpreting information from published experiments with subsystems that may be grouped into three categories: (i) the fundamental data structures that support the system as a whole; (ii) data structures that represent a computational classification of information from published papers; and (iii) the representation of neural circuitry and ethology that forms the output of the system. The design of these subsystems actively addresses the informational problems raised by my high-level strategy for studying the neural basis of behaviour.

In order to describe clearly how the design postulates a solution to the stated problem, each of NeuroScholar's subsystems will be navigated through in a relatively simplified technical format. NeuroScholar is an object-orientated system, and within this paper is presented a high-level description of the system's organization and data structures using terminology based on the symbolic nomenclature of the universal modelling language (UML). The UML is a consolidation of the work of three developers in the field of object-orientated design, J. Rumbaugh, G. Booch and I. Jacobson (Rumbaugh *et al.* 1992; Booch 1994; Jacobson *et al.* 1992), and is emerging as a widely used software engineering methodology for producing visual representations of the design of computer systems. Neuroinformatics specialists are invited to examine the project's documentation online at the NeuroScholar Web site (<http://www.neuroscholar.org>) for a full technical specification of the system in the UML. Within the text of this paper, the design of the NeuroScholar as a postulated representation for the stated problem is presented. ***Bold italics*** are used to denote the abstract representations of named classes in the NeuroScholar system.

Table 1 acts as a navigational guide for this paper, showing how the design of the main data structures (called 'classes' in standard object-orientated terminology) used in each subsystem are described in each section. All main classes are cross-indexed to their illustrative figures within this paper (with the exception of the graph theoretical representation). Table 1 describes whether specializations exist for each main class, indicating if the class forms the basis for more complex classes within NeuroScholar (for example, the 'graph' class provides a common architecture for flow charts, ethograms and connection matrices).

Table 1. A global plan of the architecture of the NeuroScholar system.

category	subsystem	main classes	fig	specializations
reusable data structures	measurement	measurement	2	yes [nominal, ordinal, interval, ratio]
		measurement_set	2	yes [nominal, ordinal, interval, ratio measurement]
		measurement_collection	2	no [but is used throughout system]
	knowledge management core	publication	3	no
		fragment	3	yes [textual, graphical]
		entity	3	yes [throughout system]
		relation	3	yes [throughout system]
		annotation	3	yes [comment, judgement, justification]
	neuroanatomical atlases	atlas	4	no
		atlas_plate	4	no
		atlas_embedded_volume	4	no
		spatial_primitive	4	yes [point, line, curve, surface, volume]
		spatial_relation	4	yes [specific to entity types]
		spatial_primitive_collection	4	no
		map	4	no
	graph theory	graph	-	yes [flowcharts, ethogram, connection matrix]
experimental model	experimental subject	research_subject	6	no
		species		no
		research_group		no
	flow charts	experiment_graph	6	no [but variable specific to domain types]
		activity_state		no
		activity_transitions		no
	data containers	branch_point		no
		experimental_manipulation	7	yes [specific to domain type]
		experimental_data		yes [specific to domain type]
		open_parameter		no [but is variable specific to domain type]
knowledge model	mapped data	mapped_observations & mapped_interventions	8	no
		mapped_histological_labeling	8, 9	no
		tract_tracing_data	10	no
		behaviour	11	no
		ethogram	11	no
	ethology	action	11	no
		action_transition	11	no
		stimulus	11	no
	stimulus neuron population	neuron_population	11, 12	yes [volume_neuronpopulation, cell_type]
		mapped_somata	12	no
		mapped_dendrites	12	no
		mapped_axons	12	no
		mapped_connectivity	12	no
		connection_matrix	13	no
	neural circuitry			

2. THE DATA MODEL

The top tier of the data model shown in table 1 describes subsystems that formulate and constrain all other aspects of the data model. Each separate subsystem will be discussed in turn. Following sections describe how the data model may represent a publication's account of experi-

mental work and the way that an individual scientist may construct a model of their own knowledge of the system.

(a) Reusable data constructs

(i) Measurements

Somewhat appropriately, the most fundamental of NeuroScholar's subsystems is concerned with the

representation of measurements. These are organized into four classes that correspond to the four measurement scales: nominal, ordinal, interval and ratio (after Stevens 1946). Nominal measurements are descriptive and based on categories (for instance, 'eye colour' is an example of a nominal measurement variable). Ordinal measurements may be ranked so that one measurement may be considered 'greater than' or 'less than' another but it is impossible to quantify by how much (for example, contestants in a beauty contest would be judged on the basis of an ordinal score). The absolute value of the difference between interval measures carries meaning, so no conclusions may be drawn about the value of the zero point (the Celsius temperature scale was an interval measure before absolute zero was discovered). Finally, ratio measurements have a well-defined zero point so that the relative sizes of different measurements may be quantified.

Classes are defined to represent individual measurements for each scale as specializations of the *measurement* class (as *nominal_measurement*, *ordinal_measurement*, *interval_measurement* and *ratio_measurement* objects) and also define the *measurement_set* class to act as a collection of *measurement* objects derived from a specified measurement scale (as *nominal_set*, *ordinal_set*, *interval_set* and *ratio_set* classes) (see figure 1). Multi-dimensional arrays of measurements are represented by the *measurement_collection* class as sets of *measurement_set* objects.

In certain situations, it may be necessary to restrict the number of possible values available to nominal measurements to a specified set of possible options. The *cv* class denotes a 'controlled vocabulary' where a limited number of possible terms are assigned specific meanings within the system. The *controlled_nominal_measurement* class is a specialization of *nominal_measurement* that has its possible values predefined as instances of the *cv* class.

(ii) *The knowledge management core*

Aspects of the system under development are not exclusive to the domain of neuroscience; rather they are general aspects of the way published knowledge may be represented and manipulated. These aspects of the theoretical framework are denoted with the term 'knowledge management core' (KMC). The primary objective of NeuroScholar is to define a 'computational intellectual environment' for neuroscientists that they can use to manipulate published information from the literature in the same way that an application programming interface (API) allows a software engineer to manipulate programming data. The system's functionality has been previously described as a set of software requirements (Burns 2001) but may be briefly described here as (i) to represent the contents of the system's target literature accurately; (ii) to permit users to interpret the literature according to their own judgement, producing a personalized representation; (iii) to provide mechanisms to allow users to merge, share and compare their individualized representations; (iv) to cross-reference the data in order to identify contradictions and discrepancies between different personalized representations; and (v) to

provide data analysis tools that help to form more powerful interpretations of the literature.

There are five fundamental classes within the KMC that underlie the organization of the whole system. These are the *publication*, *fragment*, *entity*, *relation* and *annotation* classes. Broadly speaking, the *publication* and *fragment* classes describe the contents of the literature in its original form. The *entity* class allows users to construct complex, object-orientated classifications of *fragment* instances and instances of the *relation* class form rules between pairs of instances of the *entity* class. Instances of the *annotation* class permit users to attach their comments and explanations to the other fundamental classes in an unstructured way to allow them to represent freely their ideas as users of the system. Each instance of a fundamental class is assigned a type attribute, which determines aspects of its behaviour within the system. Additionally, *entity* instances are assigned 'knowledge type' and 'domain type' attribute values. The knowledge type describes the source of the information underlying the definition of an entity instance and may take values of 'description', 'author interpretation', 'author citation' or 'user interpretation'. The domain type values describe which area of knowledge in which the *entity* instance's definition belongs and may take values of 'core' or the name of the particular class of experiments of the data and within the current design, namely, 'ethology', 'observational physiology', 'interventional physiology', 'histology' or 'tract tracing' (see §2b(ii)). Instances of the *relation* class are also constrained by knowledge type.

The design of the system is expressed as a UML class diagram in figure 2. The attributes of the *publication* class are derived from the so-called 'Dublin core' meta-dataset as a compact representation of the information required to reference a publication as an unambiguous citation (Weibel *et al.* 1995). The Dublin core meta-dataset is a simplification of the BIB-1 attribute set for use under the Z39-50 transfer protocol, which is an internationally recognized standard (ISO 1995). Each *fragment* object contains a single datum of information as it appears within the original source material, such as each sentence of the paper in the author's own words, a caption from a figure or a digitized version of a figure. As described in §1, this is a necessary feature to allow individual users to be able to reconstruct the logic (theirs and other users) in arriving at a specific interpretation, so that the accuracy and validity of the interpretations stored within the system would not have to be taken simply on trust. They could be reconstructed and thought through (and even adapted and altered) at a later date.

The *entity*, *relation* and *annotation* classes are abstract generalizations of the classes that actually implement the design by holding data. Most of the rest of this paper is concerned with specializations of the *entity* and *relation* classes. Instances of relations are association classes between two *entity* objects or between two *relation* objects (as either *interentity_relations* or *interrelation_relations*).

There are three specializations of the *annotation* class and each one fulfils a different role. Instances of the *comment* class may be attached to any instance of a fundamental class in order to express the user's opinion concerning that specific instance. Instances of

Figure 1. A class diagram showing how measurements are represented within the NeuroScholar system (see text for details).

the *justification* class record user's scientific arguments to support the definition of *entity* and *relation* instances. When *entity* or *relation* objects are defined on the basis of individual *fragment* objects, they use *frg_ent_just* or *frg_rel_just* objects (as specializations of the *justification* class) to explain the definition. Whenever *entity* or *relation* objects are defined with their knowledge type attribute set to 'user interpretation', they must be based on *entity* or *relation* objects with knowledge type attributes set to 'description' (denoting interpretations that are based on data). Of this type of knowledge-type-based justification, only the entity-to-entity association is shown in figure 2 as the *ent_from_ent* specialization of the *justification* class. Like the 'precision description codes' of the CoCoMac system, *justification* objects store an enumerated 'precision code' that denotes how strong the justification is, and, as with annotations, users are encouraged to explain the logic behind their choice in simple, understandable terms.

Instances of the *relation* class encode simple rules, with two specializations: the *interentity_relation* specifies rules that compare two instances of *entity* classes and the *interrelation_relation* class allows the comparison of two instances of the *relation* class.

Instances of the *interentity_relation* class may typically may take one of three forms: simple, spatial and set-theoretical. Every relation instance has a 'P' attribute (which is an instance of the *ratio_measure* class) indicating the probability that the relation will hold. Thus, the so-called 'simple' form of the *relation* class has a statistical component (albeit an unstructured one). This may be used to denote simple correlations (see the *stimulus_effect_correlation* association class in §2c(i)). The *spatial_relation* association class is a specialization of the *interentity_relation* association class (not shown in figure 2) and is used to relate spatial objects. These relations are based on the ORT of Stephan and Kötter (Stephan *et al.* 2000b) and may take the following

nominal values (given two spatial objects A and B): 'A occupies the same space as B', 'A is enclosed by B', 'A encloses B', 'A overlaps B', 'A is separate from, but bordering on B' and 'A and B are fully separate'. Another specialization of *relation* is the *set_relation* association class and is logically very similar to the *spatial_relation* class. This relation denotes rules between two entities that are themselves sets of components that may be related on the basis of their composition. Set-theoretical relations between two entities A and B may take the following values: 'A is equivalent to B', 'A is a subset of B', 'A is a superset of B', 'A is an overlapping set of B' or 'A and B are disjoint sets'.

When contradictions arise within NeuroScholar's multi-user environment, it may be necessary to track and resolve them. In this situation, the interaction is stored as a specialization of an *interrelation_relation* called a *contradiction* association class, which links two of the contradictory *relation* objects. This may then be linked to another specialization of the annotation class (the *judgement* class) that stores a given user's judgement in this case, favouring one relation and rejecting the other.

The *entity*, *relation* and *annotation* classes are keyed to a specific *user* object as their 'creator' attribute. Each user is assigned to one or more instances of the *group* class denoting pools of individuals that collaborate and share data. This is accomplished by the definition of an attribute called 'security level' for entities, relations and annotations. This attribute is modelled on the security model for UNIX computers where users may set a code for read/write access to the computational item for themselves as individuals, for members of the groups that they are associated with (as collaborators) and for the population as a whole. This code appears as an instance of the *cv* class. In order to keep track of how a given representation may evolve over time, a 'time-stamp' attribute is included in the description and is included in the *entity*, *relation* and *annotation* classes.

Figure 2. The ontology of the knowledge management core expressed in a simplified UML class diagram (see text for details).

(iii) *Neuroanatomical atlases*

Within the context of an object-orientated representation, a conventional brain atlas may be represented as a linked set of three classes: the *atlas* class would represent the publication as a whole; the *atlas_plate* class would represent each cross-section through the brain in question (typically corresponding to an experimental set of histological sections); and an *atlas_embedded_volume* class would represent each set of curves that fully delineate each named brain structure that appears in the atlas's nomenclature. The representation devised to hold the neuroanatomical details of published work is based on this simple concept and is illustrated in figure 3. The logical structure of the ORT used in the CoCoMac database is preserved, whilst providing the additional resource of a spatial/graphical framework for the data.

As described above, an *atlas* specializes the *publication* class, on the basis that atlases define parcellation schemes that may be used as a neuroanatomical reference by other publications. The *spatial_entity* class is a generalization of the classes used to represent volumes (the *atlas_embedded_volume* class), surfaces (the *atlas_surface* class) and curves (the *bezier_spline* class). In order to allow for uncertainty within the process of delineating brain structures, each *atlas_embedded_volume* object may be defined by up to two *atlas_surface* objects. The first surface is referred to as the 'definite boundary' that delineates the minimum extent of the volume's border and the second surface is referred to as the 'likely boundary' to delineate the maximum extent of the volume's possible border. Each *atlas_surface* object is represented by an

ordered array or stack of closed *bezier_curve* objects, each of which must lie on an atlas plate.

The computational support built into published atlases themselves will certainly develop in the near future and so the mechanisms represented here have been made extensible to accommodate other spatial constructs that are commonly used in neuroscientific datasets. Other examples of specializations of this class that are not shown in figure 3 are simple points (used in cell mapping), line segments and bezier segments (used in fibre mapping), tree-like sets of line segments (used in compartmental modelling), and sets of rectangular voxels (used in brain imaging techniques). (See NeuroScholar's current technical documentation for details of this representation at <http://www.neuroscholar.org>.)

To include the functionality of the ORT methodology, each spatial entity may be associated with *atlas_embedded_volume* objects through the use of a spatial relation (the *sp_at_vol_sptl_rln* association class). As described in §2a(ii), spatial relations simply assign a value to the association to describe how they spatially coincide (after Stephan *et al.* 2000b). The *spatial_entity_collection* class simply defines a set of spatial entities, typically an array of delineated volumes of brain tissue scattered within the spatial structure of an atlas. The *map* class forms the basis of our representation of spatial structure throughout the system and has two attributes, referring to an atlas and a 'framework' consisting of a *spatial_entity_collection* object so that arrays of data (in the form of *measurement_set* objects) may be placed into the context of an atlas.

classes

objects

Figure 3. The organization of atlas-based representations, expressed as a UML class diagram and with an illustrative example of the **atlas**, **atlas_plate** and **atlas_embedded_volume** classes (from Swanson 1998).

(iv) Graph theory

Graph-theoretical approaches provide a general mathematical representation of networks that are made up of nodes and edges. This subsystem provides an abstract, mathematical representation that may then be further specialized for specific applications. The **graph** class represents the whole graph, the **graph_node** class represents each node in the graph. Edges between nodes are represented as associations between nodes that can possess properties based on the **graph_edge** association class. The **pathway** class describes a route through the graph and the **node_set** class describes a grouping of nodes as a mathematical set. Within NeuroScholar graphs are used in three contexts: to describe the procedures

used within a scientific protocol; to represent ethograms; and to represent connection matrices.

(b) Capturing experimental evidence

The scientific objective of the current phase of this project is to build highly specified knowledge models by drawing on publicly available information from the literature. So, rather than attempting to capture the meaning of all details of papers, the design of the data model is restricted to general descriptions of experimental protocol, subjects and results. This design was based on the NeuroCore project that was formulated as part of the Brain Project of the University of Southern California, USA (Arbib & Grethe 2001). Notably, only the methods

and data are represented here, not interpretations derived from the data.

(i) *Flowcharts for expressing experimental design*

There are similarities between scientific protocols and computer programs. Both are sets of organized, constrained, sequential, branched and repeating procedural instructions. The section of the UML used to describe the flow of control with a computer system is adopted to describe procedural flow in a scientific experiment by using 'activity diagrams' to illustrate the sequential and concurrent steps of a computational procedure. Typically, activity diagrams are used to describe workflows or procedures (Rumbaugh *et al.* 1999). Essentially, activity diagrams may be thought of as formatted graphs (i.e. networks of nodes and edges) where each node plays a specified role within the figure. Thus a user could track the experience of a single experimental subject by tracing a pathway through the activity diagram. Since the whole graph represents all possible pathways, the whole diagram captures all the experiences of the experimental subjects within a single representation.

Two additional constraints should also be considered within this model. First, more than one action may be performed concurrently so that a single path should be permitted to diverge into streams (for example, a stimulus might be delivered while measurements are recorded). This is provided by the 'synchronization' event that is used to represent forking of process control within the UML. Second, the number of animals undergoing a specific transition within the graph is represented by attaching a weight to the edge that represents the transition. An example is presented below (see figure 4) to illustrate this use of activity diagrams.

As with class diagrams within the UML, activity diagrams possess a vocabulary of different diagrammatic items. The 'start state' and 'end state' for each diagram are specialized nodes within the graph (appearing as a simple filled circle and a ringed filled circle, respectively). These indicate the source node (where all edges emanate from) and the sink node (where all edges terminate) of the graph representing the flow of activity. The majority of specialized nodes are the 'activity states', which are used to denote real-world procedures that may possess substructure. The substructure of a single activity state may be described as another whole nested activity diagram (see figure 4), which is exploited in NeuroScholar to subdivide and encapsulate parts of the overall protocol. When part of a experimental protocol is repeated, an activity state is assigned to the protocol with an attached 'loop' (with the number of repetitions attached to the loop).

Activity states are linked by directional transitions (as specialized edges of the graph) that indicate what procedures follow from the current state. Two other forms of specialized node are used to denote the flow of control within the graph. Nodes that appear as diamonds are 'branch points' where the flow of the procedure diverges so that the experience of different individuals separates. This branching is usually conditional on external parameters (such as the definition of experimental groups or the values of measurements associated with other activity states). 'Synchronization' nodes appear as thick horizontal

lines in activity diagrams and act to allow the flow of activity to run in parallel across more than one pathway. This mechanism is used to separate the different components of single experimental trials. In general, the use of synchronizations defines a common space time framework for *measurement* objects, with variables that may be correlated but are not causally related. When activity states are associated with a synchronization node, they are assigned a 'start time' and an 'end time' to indicate the duration of the activity within the shared time-frame.

The set of activities that defines each individual subject's experience over the course of an experiment could then be defined as a single pathway from the 'start' to the 'end' state of the whole graph (shown in the example in figure 4 as a heavy dotted line). Different experimental groups to be represented by describing collections of multiple paths are allowed. One of the most critical pieces of information to be included in the materials and methods sections of papers is an account of the number of subjects that undergo the experimental protocol (usually stated as an '*n*'-value). Within our graph-based formulation this is represented by attaching a weight to each edge of the graph that is equal to the total number of experimental subjects involved in the specified transition.

Within NeuroScholar, the specifications described here are represented as classes that are specializations of the general classes for graphs: the *experiment_graph* class is a specialization of the *graph* class; the *start_state*, *end_state*, *synchronization*, *branch_point* and *activity_state* classes are specializations of the *graph_node* class; the loop class is a specialization of the *graph_edge* class (and transitions between states are represented by straightforward *graph_edge* objects); and *experiment_pathway* specializes *pathway* classes (to keep track of experimental subject's experience through the experiment). Each instance of the *research_subject* class is linked to an *experiment_pathway* instance to accommodate this feature. Additional associations are included to represent additional constraints, i.e. activity state nodes may contain subgraphs that represent repeatable sections of the experiment (see figure 4).

In order to illustrate these concepts, an example is presented of the use of the system to describe the first experiments elaborating classically conditioned eyelid responses in the rabbit (Schneiderman *et al.* 1962; Gormezano *et al.* 1962). This behavioural method is well established as a learning paradigm where an animal acquires an association between a conditioning stimulus (CS) (in this case, an audible tone) and an unconditioned stimulus (US) (a pulse of compressed air blown onto the cornea). The two stimuli are presented so that they coterminate (Schneiderman *et al.* 1962; Gormezano *et al.* 1962). After this 'acquisition', the animal's behaviour that would normally be triggered by the US (the movement of the nictitating membrane that protects the eye from the air puff), may be elicited simply by the presentation of the CS. It is also possible to demonstrate the 'extinction' of the learned response where the effectiveness of the CS to trigger the behaviour lessens if the CS and US are not paired within trials.

The experiment may be broken down into three nested levels of diagrams. At the top level, there are

Figure 4. Flowcharts representing the flow of organization in the original classical rabbit eyeblink conditioning experiments (from Gormezano *et al.* 1962), expressed as nested UML activity diagrams (see text for details).

simply three series of days: 'adaptation' days (where only the spontaneous movement of the eyelid was measured); 'acquisition' days (where the CS and US were paired within trials); and 'extinction days' (where the CS was presented on its own). Each experimental day is represented as an expandable UML 'activity state', with arrows denoting the transitions between states and a loop to show the number of times the activity is repeated. The organization of the experimental paradigm for an 'acquisition day' is illustrated, showing how different experimental groups are subject to different protocols through the use of a branch point. One of the blocks that form the description of individual trials has been expanded to illustrate the lowest level of organization of this process as a specialized form of activity diagram, where the start state is immediately followed by a synchronization state (indicated by a heavy horizontal bar), and the protocol diverges into three separate yet concurrent streams, two concerning the different experimental stimuli presented to the animal and another concerning the animal's behaviour. The 'end

state' of the diagram is preceded by another synchronization indicating the timed end of the experiment block.

In order to use these constructs, neuroscientific users (either the authors of papers or the collators of data into the system) will be required to construct flow charts for experiments. Although this is quite complex, it is hoped that the general design of the data structures will permit relevant papers to be incorporated straightforwardly into the model.

(ii) Data containers and open parameters

The framework designed to describe the procedural sequence of events within the NeuroScholar system has been described. At the lowest level of this description, activities such as making measurements or performing experimental manipulations require that data be attached to the activity states within the flow chart (for example, the physical parameters of the stimulus or experimental results themselves). This is accomplished by defining the **data_container** class and associating it with the **activity_state** class to link the data to parts of the experimental

of a graph, adapted from Gormezano *et al.* 1962) to capture the behaviour. These constructs allow data (in the form of **measurement_sets**) to be attached to relevant **activity_state** objects in the protocol. These methods are designed to allow users maximum flexibility in describing relevant parameters and data from experiments where methods vary between studies.

(iii) *Mapped data: interventions, observations, histology and tract-tracing data*

In cases where data must be localized within brain tissue, atlases may be used to map data by associating **measurement_set** objects with **map** objects. This is illustrated in figure 6 with an example representation of cellular labelling within the hypothalamus on level 27 of the Swanson atlas. There are four ellipsoids representing cross-sections through individual **atlas_embedded_volume** objects on the figure (forming the framework of a **map** object within this example). Each ellipsoid encloses a separate designated region of brain tissue to which we may attach a separate **measurement_set**, permitting data to be embedded into the atlas.

Mapped data are used to represent three different types of data: observational and interventional physiological data and histological data. Both the **mapped_observations** and **mapped_interventions** classes are specializations of the **experimental_data** class, allowing the **open_parameter** objects described in §2b(ii) to be used to describe a wide range of different data types. Two examples are cited of how this construct may be used to represent interventional physiological data within NeuroScholar. In an electrophysiological stimulation study of aggression, Kruk *et al.* (1983) injected current into the hypothalamus to produce aggressive responses and plotted the data over atlas plates. Similarly, in a lesion study of aggression, Canteras *et al.* (1997) managed to elicit a severe reduction of aggressive behaviour in the rat through the use of discrete cytotoxic chemical lesions within the hypothalamus, which were presented as maps within the paper's figures. As an example of the use of this construct with observational physiological data, within the same study, Canteras *et al.* also presented c-fos immunoreactive data within the hypothalamus for animals that had expressed defensive behaviour, and this could be considered as an example of a mapped observational variable.

Users of the NeuroScholar system would use the same data structures to represent the different types of data, with suitable annotations to make the differences between the representations as clear as possible. This supports the premise of the treatment of both interventional and observational physiology within NeuroScholar as simply to show that specific populations of neurons are involved in a specific behaviour, and not to attempt to describe the mechanisms of the involvement at all.

The same mapping mechanisms may be used with a more structured dataset to represent histological data. Figure 7 is a simplified class diagram showing how the **mapped_histological_labeling** class is itself a specialization of the **experimental_data** class and has two specializations of its own: the **mapped_cellular_labeling** and **mapped_fiber_labeling**. These classes use attributes with their type set to **measurement_collection** classes to

Figure 6. An illustrative example of a mapped variable, where a **measurement_collection** object is linked to a **spatial_primitive_collection** object allowing **measurement_set** objects to be embedded into an atlas. This figure could illustrate histological labelling or the positions of different classes of recording sites. (Adapted from Swanson 1998.)

represent data, and carry structure and meaning that is constrained within the system (relating to specific qualities of histological labelling localized in neural somata or axons). For the **mapped_cellular_labeling** class, the attributes relate to the cellular distribution, shape and size of the labelled cells typically using ordinal and nominal measurements. For the **mapped_fiber_labeling** class, the attributes relate to the density of varicosities and boutons along fibres in different regions, the degree of branching and the density of labelled fibres themselves.

The ultimate culmination of the descriptions of different data types is NeuroScholar's representation of tract-tracing experimental data and the experimental manipulations required to produce it. This is illustrated as a class diagram in figure 8 and makes use of the classes shown in figure 7 to represent histological data. Essentially, the data appear as two **mapped_histological_labeling** objects (associated with a single **tract_tracing_data** object): one to describe the distribution of the tracer as it was delivered into the brain at the start of the experiment; and one for the label that was taken up by neurons and transported intracellularly. The location of untransported label from tract-tracer injections is usually referred to as an 'injection site', although injections may involve multiple deposits of tracer along the injection track (requiring delineation over several separate volumes of brain tissue). This means that a mapped representation is more general and appropriate.

The representation of the methodology used within the experiment is vital for the correct interpretation of the data and is represented as a specialization of the **experimental_manipulation** class (either as a **iontophoretic_injection** or a **pressure_injection** class). The attributes of these classes take values of instances of the **measurement** class to capture the parameters of each methodology in a standardized format (the magnitude, stepping characteristics

Figure 7. Class diagram for mapped histological data within the NeuroScholar system.

and duration of the current used in iontophoretic injections; and the volume of tracer used in pressure injections).

The uptake characteristics of the tracer used within the experiment are represented within the **tracer_chemical** class, where the 'uptake_properties' attribute is an **ordinal_set** object (where the ordered values were 'none', 'weak', 'moderate' and 'strong'). For each tracer chemical, an array of four values is defined, denoting the strength of tracer uptake in the dendrites, somata, axons and terminals of a general neuronal population. Thus, if we assume that *Phaseolus vulgaris* leucoagglutinin (PHA-L) tracers have ideal properties of an 'ideal' anterograde tracer, the **tracer_chemical** array of PHA-L would be defined as dendrite: 'strong'; somata: 'strong'; axons: 'none'; terminals: 'none'. Alternatively, as a retrograde tracer, affected by uptake in fibres of passage, the **tracer_chemical** array of fluorogold might be defined as dendrite: 'none'; somata: 'none'; axons: 'moderate'; terminals: 'strong'.

(c) Knowledge models

The previous sections have been concerned with preliminary data representations that serve to support the output of the system: a comprehensive representation of the neural circuitry derived from (and transparently linked to) the literature. In this section this output representation is described and the methods used to structure it within the context of underlying data and other knowledge models. According to the design of the KMC, all instances of specializations of the **entity** class that make up a knowledge model have a 'knowledge type' set to 'user interpretation'. According to the constraints described above (and illustrated in figure 2), this requires that each component of the knowledge model must be linked (via a **justification** object) to an instance of a specialization of the **entity** class with a 'knowledge type' set to 'description'. Finally, all **entity** objects must be

linked to **fragment** objects (via **justification** objects). Thus, the knowledge model implicitly involves data and explanations that support its definition.

(i) Ethological knowledge models

The hierarchical organization of classical ethology breaks down the actions of an animal into components, based on movements of the whole animal at the highest level (but also incorporates the movements of joints, muscle groups and individual muscles; Tinbergen 1951). If the **behaviour** class defines entities that are high-level descriptions of whole classes of movement defined by their purpose (e.g. 'aggression', 'reproduction', 'classical conditioning'), then the **action** class defines entities that are sets of individual physical actions undertaken by the animal (e.g. 'biting', 'lordosis', 'blinking'). Individual actions may be grouped together to form other higher-level actions (for example, a rat performing an 'attacking' action may also be performing 'biting', 'jumping' and 'vocalization' actions). This allows us to build complex hybrid ethological entities through the use of **set_relation** classes.

Within this scheme, the **ethogram** class ties actions to behaviour. This is accomplished by describing the behavioural repertoire of an animal as a graph made up of nodes that represent individual actions and edges that represent transitions between actions. Thus, the **action** and **action_transition** classes are specializations of the **graph_node** and **graph_edge** classes, respectively. This is shown as a combined class and object diagram in figure 9. The weight of an edge in this graph represents the proportion of experimental subjects within the experiment that undergo the transition in question. Behaviour is characterized in terms of the differences between ethograms for different groups based on differences in the weights attached to the ethogram's edges. This is shown in figure 9 illustrating NeuroScholar's representation of the original experiments describing classical conditioning

Figure 8. Class diagram for tract-tracing data in the NeuroScholar system.

behaviour (Gormezano *et al.* 1962). In a way that is consistent with the literature describing this subject, the actions within the ethogram are ‘resting’ (where the displacement of the animal’s eyelid is below a certain measurement threshold), the ‘conditioned response’ (where the displacement of the animal’s eyelid above the threshold occurs before the unconditioned stimulus) and the ‘unconditioned response’ (where the displacement of the animal’s eyelid above the threshold occurs after the unconditioned stimulus). The ethogram in figure 9 shows that 93% of animals in the ‘conditioned’ group performed the transition shown when presented with a conditioned stimulus, whereas only 6% of animals underwent the same transition in the ‘control’ group. This variation of weights in the graph forms the basis of a **set_relation** object with a value of ‘disjoint’.

Importantly, within the scope of the system, no assumptions of causality are made to link stimuli directly to these behavioural transitions. The **stimulus** class has descriptive attributes (including a textual description of the population of specialized receptors that transduce the physical phenomenon into neural signals) and **stimulus** instances may be associated with a given **action_transition** via a **stimulus_effect_correlation** association class that specializes the **relation** class and forms the basis for the estimated likelihood that the stimulus is closely correlated to the transition.

In the cases where differences between ethograms may be directly correlated to the presence, absence or action of specific sets of neurons, associations may be made between **neuron_population** entities and set relations that describe the overlap or separation between different ethograms (i.e. the **eth_setRln** class, see figure 9). Importantly, if a **neuron_population** is implicated in a correlation between two **ethogram** entities that underlie the formulation of a specific behaviour, then that **neuron_population** is considered to be part of the neuronal circuitry responsible for that behaviour. These **neuron_population** objects have a ‘knowledge type’ of ‘user inter-

pretation’ and are based on data from interventional and observational physiological experiments.

(ii) *Neuron populations and neural circuits*

The **neuron_population** class forms the basis of the neural representation within the system, illustrated as a class diagram in figure 10. The **neuron_population** entity represents a generalized population of neurons and is made up of three classes: the **mapped_somata**, **mapped_dendrites** and **mapped_axons** classes. Like other so-called ‘mapped’ variables, described in previous sections, these classes define a range of attributes (involving specific measurements) across a set of spatial primitives defined within an atlas. The attributes are concerned with general features of neural architecture (such as cell or fibre density, the density of boutons, etc.), they may take quantitative or qualitative values and are expected to be refined and changed as the system’s design is tested.

The **neuron_population** class is defined to refer to ‘any population of neurons’ and has two specializations: the **volume_neuron_population** and **cell_type** classes. A **volume_neuron_population** object is defined as the population of all neurons that have their somata enclosed within a specific volume of space. In fact, connectivity studies that discuss only connections between spatially defined brain regions implicitly use these structures.

The significance of the definition of **cell_type** entities within the NeuroScholar system is that their constituent neurons should all possess the same efferent connectivity with respect to the positions of terminal fields. Consider the idealized **neuron_population** object ‘A’ in figure 10: examination of the figure shows that it consists of five cells (labelled ‘A1’, ‘A2’, ‘A3’, ‘A4’ and ‘A5’) with dendrites occupying an extended region and terminals in two separate regions in the figure (labelled R1 and R2). The diagram shows that {A1, A2, A3, A4} project to R1 and {A3, A4, A5} project to R2. If a double labelling experiment is performed where green fluorescent retrograde tracer were to be injected into R1 and red fluorescent retrograde tracer

Figure 9. Class and object diagram for the representation of the ethological component of the knowledge model (see text for details).

were injected into R2, then the {A1, A2} neurons would be labelled green, {A5} would be labelled red and {A3, A4} would be double-labelled both green and red. Thus it would be possible to define the following three **cell_type** objects: {A1, A2}, {A5} and {A3, A4} where all neurons within each **cell_type** object project to the same combination of regions. Set-theoretical rules may be used to link **neuron_population** objects together recursively in a very general way, so that the three **cell_type** objects would be related to the **neuron_population** object 'A' as subsets. A 'connection matrix' shows this in figure 10.

Connectivity is represented as a specialization of the **relation** association class acting between two **neuron_population** objects (i.e. the **mapped_connectivity** association class). This is illustrated on the right-hand side of figure 10 with two **neuron_population** objects (B and C) where the mapped terminals of C impinge over the mapped dendrites of B, demarcating a region of possible

connectivity. Note that only terminals and dendrites are discussed here, as an idealization of the actual process that involves all mapped constituents of the **neuron_population** class. The **mapped_connectivity** object attaches measurement sets that are relevant to the connectivity to that region of overlap (the density of terminals and dendrites and an indicator of whether synaptic contact is physically established; see figure 10).

There are a number of established computational techniques for analysing neural connectivity matrices including optimal set analysis, non-metric multidimensional scaling, and non-parametric cluster analysis (Young 1992; Hilgetag *et al.* 2000). A connectivity matrix is essentially a specialization of a graph, and these techniques may be encapsulated within NeuroScholar as operations of the **graph** class (illustrated in a class diagram in figure 11) and defining the **connection_matrix** class as a specialization of the **graph** class (where **connection_node** and

Figure 10. Class and object diagrams for *neuron_population* objects, illustrating how mixing of axonal targets in a *neuron_population* entity gives rise to different *cell_type* entities, and the basis for the concept of the *mapped_connectivity* association class.

connection classes are defined as specializations of the *graph_node* and *graph_edge* classes to represent *neuron_population* and *connectivity_map* objects). This provides a toolset to users of NeuroScholar that has been shown previously to allow neuroscientists to form theories based on large quantities of data that may only be collated within informatics systems such as NeuroScholar.

3. INTERACTING WITH THE DATA MODEL

Software engineering specifications usually involve descriptions of ‘use-case’ scenarios to illustrate how the system is intended to be used. This is an essential part of

the design process, forcing designers to stipulate whom the intended users are and how they can use the software to achieve specific goals. Within this section, how the functionality of the NeuroScholar system is based on the fundamental subsystems (as shown in table 1) is given.

(a) General issues of usage

The success or failure of this project will hinge on aspects of its implementation that extend beyond the scope of this paper, since only the conceptual design of the system is presented here. Several issues deserve mention. The system is sufficiently complex to require that, at the very least, users should be familiar with its design to the level described in this paper. Since the target users of NeuroScholar are

Figure 11. Class diagram showing how connection matrices may be built as a specialization of the **graph** class.

non-computational experimental scientists, this may only be accomplished with extensive online documentation. Thankfully, current trends in multimedia technology allow the presentation of audiovisual information on the Internet, so that it should be relatively straightforward to produce user-friendly documentation that explains (and even demonstrates) how to operate the system.

If the system is successful and popular, issues of bandwidth and processor overload may cause problems. Currently, an object-orientated database system (Objectivity DB) to implement data storage is being used, largely because of the scalability properties of this software. The Stanford Linear Accelerator uses Objectivity DB, which stores 169 terabytes of information at the time of writing (December 2000; <http://www.objectivity.com>).

The issues of who will maintain the database, who will mediate its contents, and who will be responsible for the process of data entry are largely unanswerable at this stage. An implementation strategy is envisaged, where the database will be maintained by the organization that hosts it. The system is designed to be inherently multi-user, so that the process of mediating the data will be restricted to ensuring that users: do not corrupt other user's data; do not exceed allocated limits of storage; and do not contravene the system's guidelines of appropriate usage. The process of data entry will probably be the rate-determining step of the usage of the system, and this issue may be addressed in three ways: by allowing end users to collate data in a highly constrained manner (and with enough users, the workload will be significantly reduced), to build links with publishers that incorporate the data entry into the publishing process, and to use semi-automated methods of information extraction from text to accelerate the entry of existing unclassified papers into the system (Hobbs 1997).

(b) Use of the knowledge management core

The usage of the knowledge management core (KMC) subsystem is relatively simple: sets of **fragment** objects (i.e. the text and figures from papers) are used to define **entity** objects with their 'knowledge type' attributes set to 'description' that classify the information content of the fragments into a constrained object-orientated representation. Consider the phrase 'tracer was injected into area X' as an example of the contents of a simple **fragment** object from a tract-tracing paper. Within NeuroScholar, this fragment would be associated to an instance of the **injection** class (specialized from the **entity** class) that

classifies the method of injections in tract-tracing experiments (see §2b for an account of the organization of classes to classify experimental descriptions). Users may then build their knowledge models from **entity** objects with their 'knowledge type' attributes set to 'user interpretation'. These **entity** objects must be linked to description **entity** objects via a **justification** annotation that briefly explains the rationale for the interpretation. Users may define **relation** objects to describe how two **entity** objects are related within specific circumstances (these will be described in more detail in the following sections).

Ideally, users will be presented with a set of full-text articles that they may read and derive definitions of **entity** objects from. Users will be required to support the definition of their **entity** objects by linking it to a set of **fragment** objects that are drawn from the original paper. Such fragments would appear in an atomic form (as either single complete sentences for text fragments, or single figures for graphics), to allow reuse by other users. Different types of entities support the process of building the structured output for the system, and the practicality of this methodology depends on how technology may be developed and leveraged to automate the process of data entry so that users can use the system straightforwardly without undue effort.

Ideally, the authors would perform this process of the classification of the descriptive contents of the paper as part of the submission process. This idea is not as unlikely as it might appear since scientific publishers are beginning to incorporate a high degree of automation in paper submission practices and are involved in the development of 'document templates' to aid this process (see Elsevier Science's 'SMARTworks' systems, <http://www.smartworks2000.com>). If the templates included a representation that was compatible with the NeuroScholar data model, this aspect of data entry could be fully automated.

(c) The use of atlases

The use of **map** objects forms the basis for the treatment of neuroanatomical detail within the system. It assigns an array of measurements to each member of a set of volumes of brain tissue in an atlas. This requires that users be capable of drawing their delineations of volumes on atlas plates. When reading papers, it is often difficult to judge the precise location of the structures in question and users are encouraged to specifically place their interpretations as accurately as they can using either 'definite

Figure 12. An illustrative example showing how an injection site would be delineated in the NeuroScholar system with an inner and an outer boundary. (a) 'Functional and anatomical organization of cardiovascular pressor and depressor sites in the lateral hypothalamus. II. Ascending projections', Allen & Cechetto (1993), reprinted by permission of Wiley-Liss, Inc., a subsidiary of Wiley. (b) Adapted from Swanson (1998).

boundary' surface entities (if they are certain of the location) or with 'likely boundary' surface entities (if they are not sure). In cases where users cannot localize the structures on the atlas plates at all, users need not provide delineations and may refer to the volume by the original name used in the publication (at which point, the publication itself should be considered an atlas).

An example of this functionality is shown in figure 12. In 1993, Allen and Cechetto performed a combined electrophysiological and tract-tracing study, where they stimulated sites in the lateral hypothalamus with injections of excitatory amino acids to elicit changes in blood pressure and then performed tract-tracing experiments by injecting tracer into stereotaxically equivalent regions in the contralateral hypothalamus (Allen & Cechetto 1993). In one case, the injection site of a tract-tracing study is described as being contained within a region referred to as the 'tuberal part of the lateral hypothalamic area' ('LHA_t') by the authors. When the accompanying figure of the injection site is viewed and delineated on a standard brain atlas plate (Swanson 1998), the region in question appears to span two atlas regions: the zona incerta (ZI) and the lateral hypothalamic area (LHA). The dense core of the injection site shown on the left hand side of figure 12 forms the basis of the delineation of the 'definite boundary', and this seems to be surrounded by a large halo of labelling that has been used as the delineation of the 'likely boundary'.

The diagram in figure 12 is unusual because the figure from the paper corresponds very well to the image on the atlas plate. In the majority of cases, the plane of section of a histological sample lies at a variable angle to the stack of atlas plates (Dashti *et al.* 1997). As is often the case when interpreting the literature, users will have to make judgements as to how they choose to interpret figures and text. NeuroScholar forces people to make their assumptions explicit by registering them into a common spatial framework and then annotating their interpretations with detailed explanations of their reasons (in several respects, this aspect of NeuroScholar is similar in concept to that described by Press *et al.* (this issue)).

4. DISCUSSION

The design of a data model from a software system is a somewhat unusual choice for the content of a neuroscientific paper but is presented here with two objectives in mind. First, it is a conceptual framework for parameters that are relevant to theoretical issues within neuroscience (in an analogy with the physical scientists, where the development of theories implicitly involves variables as the conceptualization of a situation or a problem). Second, it is an 'engineering blueprint' for developers in the field of neuroinformatics to promote interoperability between this and other neuroinformatics systems in a realistic way. In general, the most effective way of the design of a system to a user is to describe the data model (Brooks 1975).

Significantly, a forward-looking paper that appeared recently in this journal stated that 'the prime intellectual task of the future lies in constructing an appropriate theoretical framework for biology' (Brenner 1999). Such a global framework would, almost certainly, have much of its fundamental grammar expressed in the language of genetics and molecular biology, but adequate formulation of large-scale concepts at a system level would need to be developed to allow a suitable theoretical traversal of scale. The framework presented here links behavioural and neural data in a way that may be useful outside of the constrained scope of the NeuroScholar system.

The question of interoperability between neuroinformatics systems may be considered to be central to the success of the subject as a whole because, unlike the study of genetics, systems-level neuroscience has relatively few standardized components. When considering the differences between neuroinformatics systems that would need to be reconciled in order to make the systems truly interoperable, one would need to consider the fact that each system has been designed to perform a specific task, with a data model (and consequent 'world view') tailored specifically to that model with its attendant assumptions. Therefore adequate mechanisms of translation must be designed and implemented that permit data in one system to be translated to another.

However, commonalities may be based on similarities in organization. For example, if both NeuroScholar and the CoCoMac system use common mechanisms such as the ORT paradigm (see Stephan *et al.*, this issue), it may be possible to draw on common features of the design to build interoperability between the two systems. This work is currently in progress (Burns *et al.* 2001).

Other ongoing work within neuroinformatics emphasizes the consolidation of common data models to catalogue and automate shared methods (see the work concerning the common data model of Gardner *et al.* (this issue)). Additionally, large-scale resources that permit the storage of very large heterogeneous datasets such as the storage resource broker (SRB) of the National Partnership of Advanced Computing Infrastructure (NPACI) implicitly catalogue and manipulate the data models' information that is deposited there. The so-called 'metadata catalogue' (or MCAT project) provides a forum where it may be possible to define and use commonalities between data models in a coordinated way (Baru *et al.* 1998).

To sum up, this paper describes an approach to address a relatively simply stated problem: 'the comprehensive delineation of neural circuits involved in a specified behaviour from the published literature'. The solution presented in this paper is concerned with practical informational issues (e.g. how interpretations may be based on experimental descriptions and methods of data integration) as well as theoretical ones (e.g. formal representations of behavioural and neural data in a consolidated scheme).

I thank Larry Swanson, Alan Watts, Dave Lavond, Richard F. Thompson and all members of their respective laboratories for providing the environment, discussions and ideas that inspired this work. I also thank Rolf Kötter for his excellent feedback concerning this manuscript, and all the useful discussions that he and Klaas Stephan have had with me, inspiring much of the formulation of this work.

APPENDIX A: THE UNIVERSAL MODELLING LANGUAGE

An object-orientated system design is based on the definition of computational entities called 'classes' that represent categories of real-world entities (examples of UML usage are presented, based on classes required for a bibliographic database). Representations of specific members of the category are called 'objects' and these may be called an 'instance of a class' (e.g. there might be an object called 'Moby Dick', which is an instance of the 'book' class). The properties of these real-world entities are defined as a set of named 'attributes' for each class that may hold either simple or complex data as text, numbers or instances of a class (i.e. if the 'book' class possesses an 'author' attribute, the 'Moby Dick' book object could have an author 'attribute' with a string value of 'Herbert Melville'). Classes may be linked with each other via 'associations', of which there are several types.

Simple associations describe simple named relationships between two classes (the relationship between an 'author' class and a 'book' class may be modelled as a

simple association). The two ends of each association may be assigned a 'role name' describing how the class at each end contribute to the association (so that the association between books and authors would have role names 'was written by' and 'wrote' respectively). Additionally, each end of an association possesses a 'multiplicity' to describe the number of instances of each class involved in an association. An author may write one-or-more books, and there may be one-or-more authors for each book—this leads to a multiplicity of '1-to-*n*' at each end of the author–book association.

'Aggregation' associations indicate that an instance of one class is made up of a collection of instances of another (a 'book series' class might have an aggregation association with the 'book' class). A 'composition' is a more tightly constrained form of aggregation where the life-times of objects involved in the part–whole relationship are coincident. An example of this would be the relationship between 'book' and 'page' classes, since pages cannot exist without being part of a book (and a book must be made up of pages).

According to common concepts within object-orientated design, the concept of inheritance is represented as 'generalization' and 'specialization' associations. A specialization class includes all the associations and attributes of its generalized counterpart (so that a 'textbook' class is a specialization of the 'book' class and a book class is a generalization of the textbook class). For reasons dependent on the implementation methodology of the NeuroScholar system, none of the classes in the NeuroScholar data model is 'polymorphic' (i.e. all specialized classes only ever inherit from one class, although one class may be inherited by several other classes). Finally, when associations themselves possess attributes and associations, they may be modelled by an 'association class'. An example of this might describe the relationship between a book and a library, containing information about how an individual book is stored as an instance of a class. It should be clear from the examples presented above that object-orientated design methodology is a very versatile way of describing a computational representation of real-world data. Naturally, this versatility means that many alternative designs exist for any given problem.

Within the UML, the most fundamental representation of a system's architecture usually takes the form of a 'class diagram'. Within this representation, classes typically appear as tiered boxes showing the name of the class and lists of the classes' attributes and operations. Within these diagrams, associations are lines that link classes with role names and multiplicities written at each end of the line. Aggregation associations appear as a line between two classes with an open diamond appearing at the end of the class acting as a container and compositions appear the same with a solid diamond rather than an open one. Generalization associations appear as triangular-headed arrows and association classes appear as standard three-tier class boxes attached to their associations with a dashed line. For detailed discussion of all aspects of the UML see Rumbaugh *et al.* 1999 and Jacobson *et al.* 1992). Class diagrams can look complex and intimidating to new users of the UML but they present data structures in a clear, unambiguous and concisely detailed way.

REFERENCES

- Allen, G. V. & Cechetto, D. F. 1993 Functional and anatomical organization of cardiovascular pressor and depressor sites in the lateral hypothalamic area. II. Ascending projections. *J. Comp. Neurol.* **330**, 421–438.
- Arbib, M. & Grethe, J. 2001 *Computing the brain: a guide to neuroinformatics*. San Diego, CA: Academic Press.
- Baru, C., Moore, R. & Rajasekar, A. 1998 *The SDSC storage resource broker*. Toronto, Canada: CASCOS'98.
- Blacstad, T. W., Heimer, L. & Mugaini, E. 1981 Experimental neuroanatomy: general approaches and laboratory procedures. In *Neuroanatomical tract-tracing methods* (ed. L. Heimer & M. J. Robards), pp. 1–56. New York: Plenum Press.
- Bloom, F. 1978 New solutions for science communication problems needed now. *Trends Neurosci.* **1**, 1.
- Bloom, F. 1995 Neuroscience-knowledge management: slow change so far. *Trends Neurosci.* **18**, 48–49.
- Bloom, F., Young, W. & Kim, Y. 1989 *BRAIN BROWSER*. San Diego, CA: Academic Press.
- Blum, B. 1986 *Clinical information systems*. New York: Springer.
- Booch, G. 1994 *Object-oriented analysis and design with applications*, 2nd edn. Reading, MA: Addison-Wesley.
- Bowden, D. M. & Martin, R. F. 1995 NeuroNames Brain Hierarchy. *NeuroImage* **2**, 63–83.
- Brenner, S. 1999 Theoretical biology in the third millennium. *Phil. Trans. R. Soc. Lond. B* **354**, 1963–1965.
- Brooks, F. 1975 *The mythical man-month: essays on software engineering*. Reading, MA: Addison-Wesley.
- Burns, G. A. P. C. 1997 Neural connectivity in the rat: theory, methods and applications. DPhil thesis, University of Oxford.
- Burns, G. A. P. C. 1999 NeuroScholar 1.00, a neuroinformatics databasing website. *Neurocomputing* **26**, 963–970.
- Burns, G. 2001 Knowledge mechanics and the NeuroScholar project, a new approach to neuroscientific theory. In *Computing the brain: a guide to neuroinformatics* (ed. M. Arbib & J. Grethe). San Diego, CA: Academic Press.
- Burns, G. A. P. C. & Young, M. P. 2000 Analysis of the connectional organization of neural systems associated with the hippocampus in rats. *Phil. Trans. R. Soc. Lond. B* **335**, 55–70.
- Burns, G., Stephan, K., Ludaescher, B., Gupta, A. & Kötter, R. 2001 Towards a federated neuroscientific knowledge management system using brain atlases. *Neurocomputing* **38–40**, 1633–1641.
- Canteras, N., Chiavegatto, S., Valle, L. & Swanson, L. 1997 Severe reduction of rat defensive behavior to a predator by discrete hypothalamic chemical lesions. *Brain Res. Bull.* **44**, 297–305.
- Canteras, N. & Swanson, L. 1992 The dorsal premammillary nucleus: an unusual component of the mammillary body. *Proc. Natl Acad. Sci. USA* **89**, 10 089–10 093.
- Dashti, A., Ghandeharizadeh, S., Stone, J., Swanson, L. & Thompson, R. 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. *Cerebr. Cortex* **1**, 1–47.
- Gerfen, C. R. & Sawchenko, P. E. 1984 An anterograde neuroanatomical tracing method that shows detailed morphology of neurons, their axons and terminals—immunohistochemical localization of an axonally transported plant lectin, *Phaseolus vulgaris* leucoagglutinin (PHA-L). *Brain Res.* **290**, 219–238.
- Gormezano, I., Schneiderman, D. E. & Fuentes, I. 1962 Nictating membrane: classical conditioning and extinction in the albino rabbit. *Science* **138**, 33–34.
- Grant, E. C. 1963 An analysis of the social behavior of the male laboratory rat. *Behavior* **21**, 260–281.
- Hastings, M. H., Winn, P. & Dunnett, S. B. 1985 Neurotoxic amino acid lesions of the lateral hypothalamus: a parametric comparison of the effects of ibotenate, N-methyl-D,L-aspartate and quisqualate in the rat. *Brain Res.* **360**, 248–256.
- Hilgetag, C. C., Neill, M. O. & Young, M. P. 1996 Indeterminate organization of the visual system. *Science* **271**, 776–777.
- Hilgetag, C., Burns, G., O'Neill, M., Scannell, J. & Young, M. 2000 Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Phil. Trans. R. Soc. Lond. B* **335**, 92–110.
- Hobbes, J. R. 1997 FASTUS: a cascaded finite-state transducer for extracting information from natural language text. In *Finite state language processing* (ed. E. Roche & Y. Schabes), pp. 383–406. Cambridge, MA: MIT Press.
- Hubel, D. H. & Wiesel, T. N. 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154.
- Hunt, S. P., Pini, A. & Evan, G. 1987 Induction of c-fos-like protein in spinal cord neurons following sensory stimulation. *Nature* **328**, 632–634.
- ISO 1995 *ISO 23950 Information retrieval (Z39.50)*. Bethesda, MD: NISO Press.
- Jacobson, I., Christerson, M., Jonsson, P. & Overgaard, G. 1992 *Object-oriented software engineering: a use case driven approach*. Reading, MA: Addison-Wesley.
- Kobbert, C., Apps, R., Bechmann, I., Lanciego, J., Mey, J. & Thanos, S. 2000 Current concepts in neuroanatomical tracing. *Prog. Neurobiol.* **62**, 327–351.
- Kollack-Walker, S., Watson, S. J. & Akil, H. 1997 Social stress in hamsters: defeat activates specific neurocircuits within the brain. *J. Neurosci.* **17**, 8842–8855.
- Kristensson, K. & Olsson, Y. 1971 Retrograde axonal transport of protein. *Brain Res.* **29**, 363–365.
- Kruk, M. R., Van der Poel, A. M., Meelis, W., Hermans, J., Mostert, P. G., Mos, J. & Lohman, A. H. 1983 Discriminant analysis of the localization of aggression-inducing electrode placements in the hypothalamus of male rats. *Brain Res.* **260**, 61–79.
- Lammers, J. H., Kruk, M. R., Meelis, W. & Van der Poel, A. M. 1988 Hypothalamic substrates for brain stimulation-induced attack, teeth-chattering and social grooming in the rat. *Brain Res.* **449**, 311–327.
- Lasek, R., Joseph, B. S. & Whitlock, D. G. 1968 Evaluation of a radioautographic neuroanatomical tracing method. *Brain Res.* **8**, 319–336.
- Morton, B., Blanchard, R. J., Lee, E., Hanchano, D., Cabebe, L. & Blanchard, D. C. 1984 The use of [14C]-2-deoxyglucose to detect regional brain activities with aggressive and defensive behaviors. In *Biological perspectives on aggression* (ed. K. J. Flannelly, R. J. Blanchard & D. C. Blanchard), pp. 295–304. New York: Liss.
- Nicolelis, M., Tinone, G., Sameshima, K., Timo-Iaria, C., Yu, C. & Van de Bilt, M. 1990 Connection, a microcomputer program for storing and analyzing structural properties of neural circuits. *Comput. Biomed. Res.* **23**, 64–81.
- Nicolelis, M. A., Ghazanfar, A. A., Fagg, B. M., Votaw, S. & Oliveira, L. M. 1997 Reconstructing the engram: simultaneous, multisite, many single neuron recordings. *Neuron* **18**, 529–537.
- Paxinos, G. & Watson, C. 1982 *The rat brain in stereotaxic coordinates*. London: Academic Press.
- Paxinos, G. & Watson, C. 1986 *The rat brain in stereotaxic coordinates*. London: Academic Press.
- Risold, P., Canteras, N. & Swanson, L. 1994 Organization of projections from the anterior hypothalamic nucleus: a *Phaseolus vulgaris*-leucoagglutinin study in the rat. *J. Comp. Neurol.* **348**, 1–40.

- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. & Lorenson, W. 1992 *Object-oriented modeling and design*. Englewood Cliffs, NJ: Prentice-Hall.
- Rumbaugh, J., Jacobson, I. & Booch, G. 1999 *The unified modeling language reference manual*. Reading, MA: Addison-Wesley.
- 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., Benson, P. J., Tóvée, M. J., Blakemore, C. & Young, M. P. 1996 Visual motion processing in anterior ectosylvian sulcus of the cat. *J. Neurophysiol.* **76**, 895–907.
- Scannell, J. W., Burns, G. A. P. C., Hilgetag, C. C., O’Neil, M. A & Young, M. P. 1999 The connectional organization of the cortico-thalamic system of the cat. *Cerebr. Cortex* **9**, 277–299.
- Schneiderman, N., Fuentes, I. & Gormezano, I. 1962 Acquisition and extinction of the classically conditioned eyelid response in the albino rabbit. *Science* **136**, 650–652.
- Shahabi, C., Dashti, A. E., Burns, G. A. P. C., Ghandeharizadeh, S., Jiang, N. & Swanson, L. W. 1999 Visualization of spatial neuroanatomical data. In *Proc. Visual’99 Conf.*, Amsterdam, pp. 801–808. Springer LNCS volume 1614. Springer.
- Sheng, M. & Greenberg, M. E. 1990 The regulation and function of *c-fos* and other immediate early genes in the nervous system. *Neuron* **4**, 477–485.
- Skinner, J. E. & Lindsley, D. B. 1968 Reversible cryogenic blockade of neural function in the brain of unrestrained animals. *Science* **161**, 595–597.
- Stephan, K. E., Hilgetag, C. C., Burns, G. A. P. C., O’Neill, M. A., Young, M. P. & Kötter, R. 2000a Computational analysis of functional connectivity between areas of primate cerebral cortex. *Phil. Trans. R. Soc. Lond.* **B355**, 111–126.
- Stephan, K. E., Zilles, K. & Kötter, R. 2000b Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Phil. Trans. R. Soc. Lond.* **B355**, 37–54.
- Stevens, S. 1946 On the theory of scales of measurement. *Science* **103**, 677–680.
- Swanson, L. 1995 Mapping the human brain: past, present, and future. *Trends Neurosci.* **18**, 471–474.
- Swanson, L. W. 1998 *Brain maps: structure of the rat brain*. Amsterdam: Elsevier.
- Tinbergen, N. 1951 *The study of instinct*. Oxford University Press.
- Weibel, S., Godby, J., Miller, E. & Daniel, R. 1995 Workshop position paper. OCLC/NCSA Metadata Workshop Report, Dublin, Ohio.
- Weiss, P. 1941 Self-differentiation of the basic patterns of coordination. *Comp. Psychol. Monogr.* **17**, 1–96.
- Wilder, B. 1896 Neural terms, international and national. *J. Comp. Neurol.* **6**, 216–352.
- 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.* **B252**, 13–18.