CHAPTER 6.3

Knowledge Mechanics and the Neuroscholar Project: A New Approach to Neuroscientific Theory

Gully A. P. C. Burns
University of Southern California Brain Project, University of Southern California, Los Angeles, California

6.3.1 An Introduction to Knowledge Mechanics

In any scientific discipline, the role of the published literature is to provide an "intellectual environment" for a domain of knowledge where new ideas are presented, contradictions and controversy may be formally aired, and experimental results and their interpreted meanings are archived for subsequent retrieval by other workers. Individual scientists interact with this environment by writing and reading papers. Recent technological developments have accelerated the speed at which this interaction may take place, in terms of the delivery of published information to scientists and of their submissions to the literature, but at present few tools have been developed to help them manage and understand this ever-increasing body of information.

The neuroscience literature has intrinsic complications due to the fact that its subject matter is both broad and deep. It involves many different scientific subdisciplines ranging from animal behavior and psychology through cellular anatomy and physiology to studies of molecular biophysics and biochemistry. Each of these subdisciplines involves different types of data and different conceptual approaches. The scholarly basis of a given field usually involves an enormous amount of in-depth information. For example, there are over 500 different brain regions in the rat (Swanson, 1998), each one of which may be made up of approximately five cell groups, which may, in turn, project to ten other cell groups. This means that the number of connections defining the circuitry between cell groups is probably of the order of 25,000 different macroconnections (see Chapter 4.1). No single individual, no matter how great a scholar, can hope to incorporate anything more than a subset of the available information into his or her thinking without some form of computational support.

A trend of research began in the early part of the 1990s that was concerned with the problem of summarizing the contents of large numbers of papers in a computational format so the summaries could then be analyzed with mathematical methods. These studies were concerned with the global organization of neuroanatomical circuits in the brain and used formal approaches to store the interpretations of the researchers who were collating data from the literature (Burns, 1997; Felleman and van Essen, 1991; Scannell *et al.,* 1995, 1999; Stephan *et al.* 2000; Young, 1993). The practical

experience of performing these studies highlighted the main practical obstacles of this approach. First, the process of summarizing the literature is extremely time consuming and requires expert knowledge. Second, no formal methods exist for making summaries forwardly compatible with subsequent work by other collators. Finally, these studies were limited to the neuroanatomical tract-tracing literature and were based on a data model that was based on a simplification of neurobiological concepts present in the primary literature (i.e., descriptions of neuronal connections did not permit axon collaterals to be represented).

The NeuroScholar project represents a progression of this technology into a new phase by introducing key concepts from the field of informatics. Consider "data" as the lowest level of known facts; "information" as data that have been sorted, analyzed, and interpreted; and "knowledge" as information that has been placed in context of other known information (Blum, 1986). If we apply these concepts to the contents of a given research publication, each individual experimental observation or fact may be represented as "data," the results and methods sections can be considered to be a conglomeration of data and therefore comprise the publication's "information," and the introduction and conclusions comprise the publication's "knowledge." Within the system described here, we will also consider the interpretations of the readers of the paper to be knowledge that may be potentially informative. Additionally, there is no reason to assume *a priori* that different collators will interpret the literature in the same way; therefore, the practicalities of solving the stated problem require a large-scale, multi-user, knowledge-base management system that can represent and contrast multiple interpretations of neuroanatomical data present in the literature.

By defining a rigorous computational approach to implementing these concepts, we seek to define a utilitarian theoretical framework for published neuroscientific data, information, and knowledge as a computational tool. Not only should this framework capture and represent the relevant information, but it should also synthesize it into predictive theories that can be used both by experimentalists and theorists. Throughout this chapter, we will refer to this framework as "Knowledge Mechanics," alluding to the way in which "Quantum Mechanics" or "Statistical Mechanics" defines a useful theoretical framework for certain physical systems. The long-term objective of NeuroScholar as a system that implements a knowledge mechanical paradigm 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. Although the work to date on the NeuroScholar project has emphasized the representation and manipulation of neural connectivity data, the knowledge mechanics framework has been designed and implemented as a general software solution. This will be discussed in more detail in the section concerned with software design.

This chapter will seek to describe how Knowledge Mechanics could achieve these goals and will examine the philosophical basis of the concept of "theory" in neuroscience and how a knowledge mechanical approach may address key issues that would otherwise be impossible to challenge. A discussion about the software requirements of NeuroScholar is followed by a description of the main design features of our data model with worked examples.

6.3.2 Concept of "Theory" in Neuroscience

The definitions of the word "theory" that are appropriate to scientific application of the word are listed below (Merriam-Webster, 2000):

1. "The analysis of a set of facts in their relation to one another"
2. "The general or abstract principles of a body of fact, a science, or an art"
3. "A plausible or scientifically acceptable general principle or body of principles offered to explain phenomena (e.g., wave theory)"
4. "A body of theorems presenting a concise systematic view of a subject (e.g., theory of equations)"

These definitions rely on the concept that a theory is some form of explanation of a set of data and among the listed definitions there is a progression of reliability and rigidity as one descends the list. This graded measure of reliability can be defined in terms of the synonyms "hypothesis," "theory," and "law." A hypothesis does not provide anything more than a tentative explanation of a given phenomenon, whereas a theory implies a more tightly constrained explanation backed up by more supporting evidence. A law suggests a carefully defined predictive statement to which there would be few, if any, exceptions (Merriam-Webster, 2000).

The goal of scientists should ultimately be to progress from hypotheses to laws in their explanations of nature's phenomena. These explanations must be made in a way that is appropriate to the state of our knowledge of the subject in question so that rigidly defined laws are only applied in situations where enough experimental evidence can be produced to support them. If the practical goal of scientific theory is to explain the mechanisms that give rise to existing data and make predictions that can be tested experimentally, then this project's objective should be to provide users with a way of making testable predictions for experiments.

Within the discipline of neuroscience, we postulate that hypotheses are put forward in the conclusions sections of papers as possible explanations of specific experimental results. When considered *en masse*, these

hypotheses represent the theoretical framework of the subject and are often expressed in review articles or textbooks. Different hypotheses often contradict each other, and the process of constructing a coherent theoretical "story" (as these things are often described) from this mosaic of different ideas is difficult. We impose the following practical guidelines for theories derived from the literature under the knowledge mechanics framework:

1. The theory should take as much as possible of the data pertaining to a system, behavior, or phenomenon into account.
2. The theory's limitations, including any supporting assumptions, should be stated clearly so that when the theory encounters anomalous data the assumptions may be tested to see if an alternative assumption could accommodate the anomalous data.
3. The degree of abstraction of the theory should not simplify the essential details of existing data. Unfortunately, much of the data in neuroscience are so complex that some simplification is necessary and judgment must be employed to decide which aspects of the data are "essential" and which are not.
4. Any specialized terminology should be defined unambiguously. This is important, as commonly used words (such as the word "connection," for example) may have more than one meaning. Some aspects of nomenclature carry very different meanings depending on the source, meaning that often it is essential to qualify the use of an expression according to a cited definition of a third party.
5. The logical argument of the theory should be self-contained, transparent, and fully explained, so nothing need be taken on faith from the theory's originator.
6. Theories should provide predictions that may be tested in addition to sensible explanations of existing data. This not only provides a validation of the theory but also gives the theory value in terms of its usefulness to the scientific process.

Although these requirements might seem somewhat obvious and based on straightforward common sense, the task of actually implementing them is extremely difficult for the following reasons.

1. The literature is too huge for single individuals to search exhaustively.
2. The act of linking highly heterogeneous and disparate facts together to form theories is very difficult to achieve computationally, and the vast majority of neuroscientists use interpretative reasoning rather than mathematics to solve their problems.
3. Some of the information in the literature is inaccurate.
4. Neuroscientific data are often qualitative and are not measured according to standard units.
5. Functional explanations of neural phenomena are often vague and are usually stated with some qualifying doubt. This is because many of the fundamental questions concerning these explanations have simply not been answered so that many important concepts cannot be defined precisely (e.g., "neural information," "signals," "pathway," "system," "organization").

Neuroscientists often express their ideas, concepts, and models in the form of written prose. This could be considered the lowest common denominator in terms of communicating precise ideas about data. Consequently, the way in which information is transferred between papers is dependent on the language used to express the information which can lead to problems of semantics, vocabulary, and nomenclature; for example, the problems surrounding neuroanatomical nomenclature can be extremely restrictive and confusing (Swanson, 1998).

The remainder of the chapter will be involved with how the NeuroScholar system seeks to solve these problems.

6.3.3 High-Level Software Requirements and Fundamental Design Concepts of the NeuroScholar System

The general issues raised in the last section may be addressed more fruitfully if we consider a specific example: previous work in building databases for summarizing neuroanatomical connection data from the literature (Burns, 1997; Burns and Young, 2000) highlighted several design principles that should be considered carefully in the design of the next generation of solutions to this problem. These issues are listed below:

1. The task of collating information from experimental neuroscientific papers into a machine-readable format is extremely time consuming and should be performed by expert scientists in the field. Even then, there will be differences in interpretation between different collators.
2. It is not sufficient to read a paper, to classify and summarize it, and then only record the summary. It is essential to record the reasoning that led collators to their chosen classification so that subsequent users transparently understand the basis for the chosen summary. Thus, an important design feature for any computational classification of the literature is that interpretations of data should be easy to follow, either by providing the data that underlie the interpretation or by explicitly explaining the interpretation.
3. The way that data are selected as "reliable" or rejected as "unreliable" should be made explicit.
4. The way in which the data are represented computationally should be as conceptually close to the intended representation of the publication's author as possible. I feel that we have to place trust in the experimentalists producing the data and to stay as faithful as possible to the way they see their data.

5. Translating between nomenclatures is of fundamental importance and may be formally addressed by encoding a computational method of translation (Stephan *et al.,* 2000b).

The translation of these design principles into a real application is an exercise in software engineering, requiring at least two separate and clearly written components: a list of the requirements of the software and a description of the software's design. This section will discuss the fundamental, general aspects of the requirements and design, and more detail will be provided in following sections.

At the time of this writing, the NeuroScholar system is a work in progress. A small-scale demonstration is implemented on the World Wide Web and may be accessed via http://neuroscholar.usc.edu/. This chapter is a presentation of software requirements and design principles with the understanding that very few of the concepts presented here have been implemented as an online service at the present time.

Software Requirements

In order to appreciate the design requirements of the system, consider the following worked example. Let us imagine the scene at a neuroscientific laboratory meeting in the future where three users of the NeuroScholar system are discussing aspects of neuronal circuitry. The discussion concerns how two regions of the brain interact with each other, and there is some disagreement. For illustrative purposes, let's call the users Alice, Bob, and Colin.

Alice is an expert on the connections of one of the regions in question (region A). Similarly, Bob is an expert on the second region (region B), and Colin is an expert on the electrophysiology of both regions under certain behavioral paradigms. All have used NeuroScholar to structure their consideration of the literature. Alice and Bob concur about the majority of pathways between the two regions but disagree about some of the connections. Colin knows much less than his two colleagues about the connections but would like to use their expertise to plan his next experiment. In this situation, the NeuroScholar system would allow them to perform the following tasks:

1. If Alice describes all the outputs of region A and Bob describes the inputs of region B, their accounts could be joined by searching for overlap between the two schemes. NeuroScholar could identify any causes of disagreement or differences of interpretation. Alice and Bob could then reconcile any controversy by looking at the reasons why each of them chose the interpretations that they did by examining the annotations they made when reading and re-reading the original papers and by refreshing their understanding of important points in primary literature. Thus, they could then put forward a unified scheme of connections in which they both believe and which describes the pathways from region A to region B.
2. Alice and Bob might then search the global NeuroScholar database for any information that they might have left out of their individual descriptions. If they disagree with interpretations of other users, then they can comment on the other users' interpretations and make their own.
3. They could analyze this scheme, either intuitively or with multivariate statistics, to look for hitherto-unknown high-level organizational principles.
4. They can make this scheme or high-level interpretations available to Colin, who would add his accounts of the electrophysiological properties of the regions in question. It would be immediately apparent to Colin where gaps in the literature revealed lack of data. He could attempt to synthesize his electrophysiological scheme into the scheme already provided by the anatomists or point out where the anatomist's ideas did not mesh with the electrophysiological data.

This example is only illustrative (descriptions of how some of these software requirements may be fulfilled will follow in later sections of this chapter), but it describes elements of the practical requirements made on the NeuroScholar system. More technically, the functionality of the system is based on the concept that users may build a computational representation of their knowledge, which they may then use in several different ways. This functionality is broken down into its component parts below.

Building the Representation

The process of building a representation of users' knowledge in NeuroScholar is accomplished by placing a computational structure on the data that adequately captures the concepts of neuroscience theory. This is equivalent to devising a suitable ontology or data model for it (an ontology may be described as "descriptions of the domain knowledge of some field;" see van Heijst *et al.,* 1997). This ontology must be able to capture as much relevant information as possible, including low-level data from experimental results and methods, high-level interpretations from within the literature, and comments, observations, judgments, and interpretations from individual users of the system. The ontology should be well defined enough to be able to formalize any common aspects of papers from different experimental protocols, but it must also be extensible to allow non-standard, original data to be included.

The adage that "knowledge is power" is especially true in the context of informatics. NeuroScholar is directly concerned with the storage and manipulation of users' expert opinions concerning their colleagues' published work. Thus, the contents of the knowledge base carry political significance, and care must be taken concerning the security of the system. Users may choose to publish

their knowledge model within the system and to restrict access to specified users or everyone, or to maintain absolute privacy, sharing data with no one.

Adjusting the Representation

Just as an individual's interpretations will change over time, individual users' knowledge models will need to be updated, adjusted, and maybe even completely reinvented. In order to keep track of all changes made to the database, detailed access logs will be maintained. Users will be encouraged to timestamp "versions" of their knowledge representation based on their particular outlook at a particular time.

Querying the Representation

The retrieval of knowledge from the system may rely on the structured, interconnected nature of the ontology to give the user the capability to query data, information, or knowledge in a combinatorial way. At present, no plans are being implemented to translate language-based queries into the system, so users would require a certain degree of familiarity with the data model in order to construct their queries appropriately. However, a well-designed user interface will provide much of the desired functionality.

Users will be able to tailor their questions in terms of the type of experiments in which they are interested (e.g., "show me all the available injection sites from tract-tracing studies using horseradish peroxidase") or in terms of a specific interpretation (e.g., "show me all the inputs to the supramammillary nucleus in the rat") or even in terms of the interpretations of data (e.g., "show me M. A. Arbib's interpretations of the connections reported in Swanson *et al.,* 1997"). Queries will be made combining data from different experiments (e.g., "show me the areas in the brain that contain neurons that project to the lateral hypothalamic area and also express ranatensin" or "what is the distribution of all peptides that are expressed in the brain of the rat when it is performing hunting behavior?").

Validating the Representation

The system may automatically search for contradictions between users or publications in order to highlight areas of controversy. When a specific contradiction is found, the system informs relevant users and allows them to make a judgment about the contradicting options. In this circumstance, users will be required to explain any judgments they make in order to make the logic behind their decision as transparent as possible for other users.

Analyzing the Representation

In some types of experiment it will be possible to perform "interpretive analyses" where information from the results section of the paper may be interpreted by the system according to logical parameters that would be stipulated as part of the definition of the experiment type. This will be discussed in more detail in the "Software Design" section of this chapter.

Users' knowledge representation could be compiled into a global summary that represents their understanding of a specific phenomenon or system. This summary could provide the input to secondary analyses either for visualization (as was the case for the neuroanatomical connection databases that were the forerunners of NeuroScholar; see Young, 1992), or to investigate organizational properties of the data. Techniques such as non-metric multidimensional analysis (NMDS), non-parametric cluster analysis (MCLUS), and optimal set analysis (OSA) have been successfully applied to similar problems in the past. Analyses such as these may help expert users interpret their data to make experimental predictions. Scannell *et al.* (1995) used NMDS to analyze a database of connections to successfully predict the presence of plaid-pattern-sensitive cells in the anterior ectosylvian sulcus of the cat. They subsequently identified these cells electrophysiologically as predicted (Scannell *et al.,* 1996). It must be stressed that these approaches are simply tools to aid the intuitive powers of neuroscientists and that we are not claiming that these approaches will circumvent the role of human reasoning in neuroscience.

Statistical approaches may also be used to generate accurate summaries by searching for optimal solutions that satisfy the many well-defined constraints that exist in the data. For example, consider a situation where the act of comparing data that originate from two different experiments is unreliable, but comparisons between data points from within the same experiment are reliable. In cases such as these, it would be possible to compile a set of meta-data comprising all the relevant constraints that could then be analyzed globally to produce global constraints for the whole system. Methods like these were used to calculate finely graded connection weights for the rat visual system (Burns *et al.,* 1996). This was accomplished on the assumption that the density of labeling in tract-tracing studies is correlated with the anatomical strength of the connection, but different tracer chemicals have different sensitivities. Thus, comparisons of labeling density within the same experiment (or tentatively between experiments that use the same technique) reflected differences in connection strength that could not be inferred from comparisons between experiments. This general approach was also used to convert between parcellation schemes in macaque monkey (Stephan *et al.,* 2000b).

Inter-representation Communication: Importing and Exporting

The apparently simple task of collating information from the literature into some form of repository is too time consuming for a single individual to tackle.

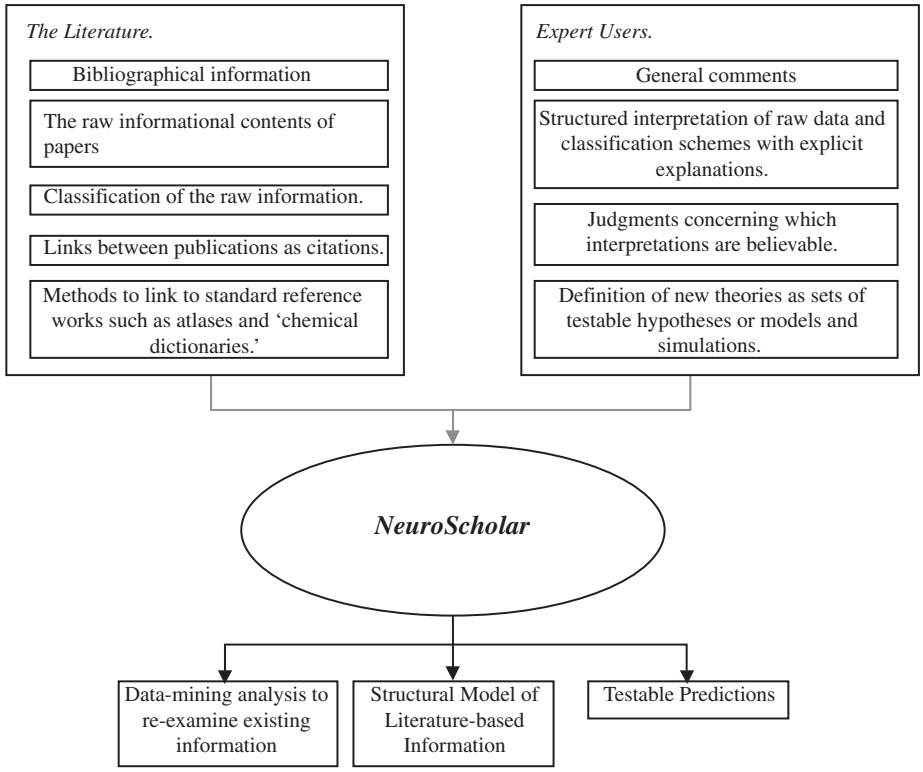

Figure 1 High-level inputs and outputs of NeuroScholar.

Consequently, NeuroScholar has been designed to be a multi-user system in order to distribute the workload of data entry effectively. This will mean that data entered by any given user may be made available to other users through conventional database security methods.

High-Level Inputs and Outputs

The high-level functionality of the system may also be described in terms of data flow, illustrated in Fig. 1.

Software Design

The data structure of our ontology for published information is probably the most important single feature of the system. We implement a hierarchical scheme for data structures in NeuroScholar which we here propose as a possible conceptual framework for "Knowledge Mechanics" as an example of a knowledge management system to fulfill our software requirements. This scheme is based on relational database design and is illustrated in Fig. 2. The highest level structures are referred to as *views*, which are made up of *primitives*. The different types of *primitives* are illustrated below in Fig. 3 and are themselves made up of *tables* (which, in turn, are made up of *tableColumns*). *Tables* and *tableColumns* are defined in terms of the computational implementation of the system and will not be considered in detail here.

Views are equivalent to the high-level representations of primitives and their links to other primitives. For

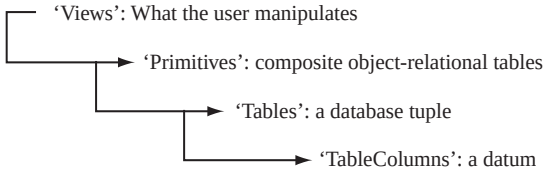

Figure 2 Global hierarchy of data structures in NeuroScholar.

example, if a journal article has three authors, a view that represents that article contains a *journalArticle* primitive and three *person* primitives. The types of primitive that we define within the knowledge mechanics paradigm are *publications*, *fragment*, *objects*, *properties*, *relations*, and *annotations* (where the *annotations* have been separated into three types: *comments*, *justifications*, and *judgements*). Each primitive is responsible for a different function within in the system. The associations between primitives are shown in Fig. 3 as a class diagram using the universal modeling language notation (UML; see Rational Software Corp., 1997). The use of UML in database design provides a powerful, versatile way of communicating schema design (see Muller, 1999, for an excellent introduction to this methodology).

We describe the following *primitives* within our system: *publications*, *fragments*, *objects*, *properties*, *relations,* and *annotations*. There are two subtypes of objects: *interpretationObjects* and *citationObjects*. *Annotations* are themselves an abstract class (i.e., every annotation must belong to one of three subtypes: *comments*, *justifications,*

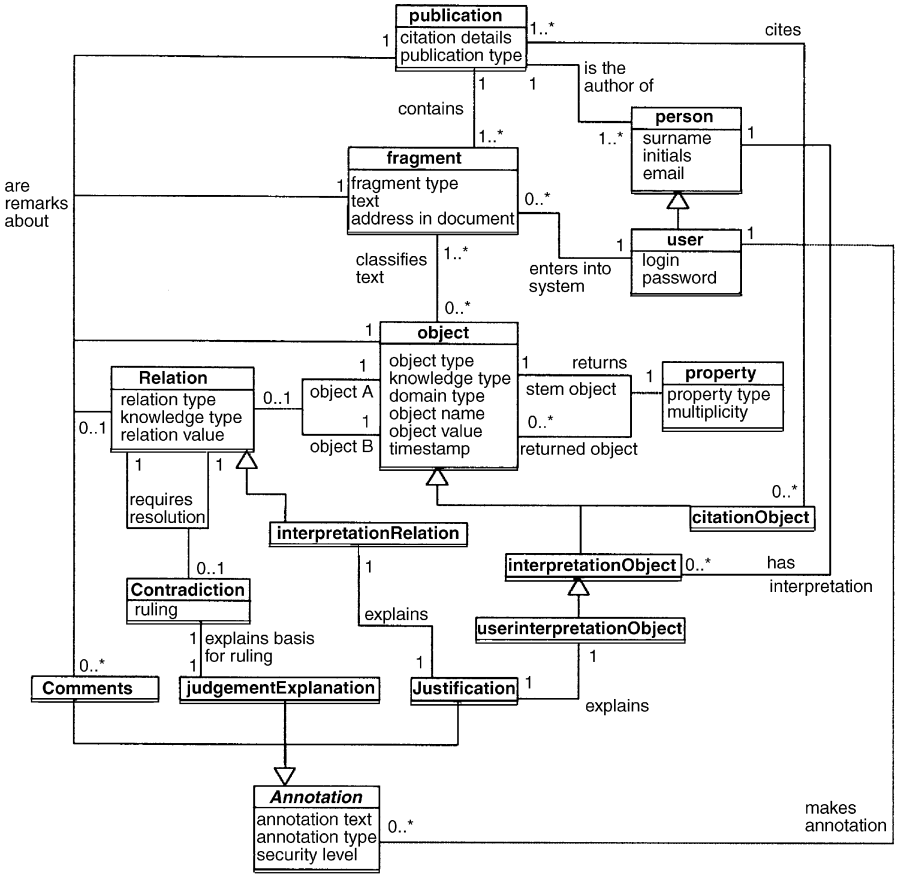

Figure 3 The ontology of the Knowledge Mechanics core expressed in a simplified universal modeling language (UML) class diagram (see text for details).

and *judgmentExplanations*). These inheritance relationships are represented as triangular-headed arrows ("generalization associations" in the UML) in Fig. 3. In the interest of presenting a concise figure, we abbreviated the schema in Fig. 3 by omitting the mechanism of how types are constrained in order to ensure that fragments, objects, their properties and relations, and annotations fit together in a highly structured way. UML representations of methods have also been omitted from the diagram.

Publications Themselves Representing the Author's Own Words

One of the problems encountered in designing a database system for neural connectivity was that, when I returned to my interpretations after some time had passed, I could no longer remember the original reasoning I had used to generate them. I then had to reread the original text to reconstruct my argument which led me to the realization that it was necessary to store both the individual phrases that were responsible for a given interpretation and their precise location in the text in the database itself. It should be noted that this requirement has legal implications, as it may contravene aspects of copyright law that control access to the content of commercial publications. This is a vital design feature and could be performed by storing links to published material in online journals, if necessary. *Publication* primitives contain the details of the source of the information being represented in the system, corresponding to instructions on how to obtain a copy of the source from outside the system. *Fragment* primitives contain a single datum of information as it appears within the original source material.

Constructing Complex Objects for Neuroscience

The role that is played by *object* primitives in NeuroScholar is symbolized by their central position in the schema shown in Fig. 3. We use *objects* as the representations of neuroscientific concepts, and we define the interrelationships between these concepts by defining properties and relations for them.

Object primitives contain the classification of a concept and are classified according to "domain type" and "knowledge type." Domain types are based on a classification of the subject under consideration; in NeuroScholar, this is defined by the experimental method being used to obtain the results that support whatever classification is being represented by the object in question.

For example, this means that we differentiate between objects defined from "tract-tracing studies" and "electrophysiological studies." Knowledge types are based on informatics-based considerations of the classification being used for the object so that we differentiate between an object that is based on primary descriptions of experimental results and an object that is based on a citation of another author's interpretations of the results. Knowledge types in use in the current design of the system are *author citation* (which requires links to a *publication* primitive), *interpretation* (with subtypes *author interpretation* and *user interpretation)*, *description*, and *calculation* (the representation of interpretive analyses described above). Domain types in use in the current design of the system are *anatomical experiments*, *core objects*, *physiological recording experiment*, *physiological stimulation experiment*, *tract-tracing experiment*, and *tract-tracing method experiment*.

Within the framework of this system, *properties* and *relations* are *primitives* that operate on *objects*. A given *property* acts on an *object* to return a second *object*, thus enabling the systems' designers to utilize an object-oriented-like approach to classifying information and knowledge. The cardinality of these relations (i.e., how many *objects* should be returned by a given *property* when acting on a given *object* and how many *objects* return a given *object* when acted on by a given *property*) may be defined explicitly, allowing the construction of large numbers of interconnected, versatile, complex data constructs (for a full description of object-oriented design, see Pressman, 1992). *Relations* may be used to attach typed values to pairs of *objects*. This enables users to compare and link *objects* according to structured rule sets that represent characteristics of the information or knowledge in use. Thus, users may describe potential overlap between parcellation schemes (Stephan *et al.,* 2000b), compare the strength of neuroanatomical connections (Burns *et al.,* 1996), highlight contradictions between different classifications of the same data, or define any number of rule-based observations concerning two *objects.*

Fig. 4 illustrates the most important composite object in NeuroScholar: a so-called *neuronPopulation (interpretation, core)* object. This defines a generalized population of neurons and is made up of *somata(interpretation, core)*, *dendrites(interpretation, core)*, *axons (interpretation, core),* and *terminalField(interpretation, core)* subobjects which contain relevant data describing the characteristics of those components. A type of object of importance in this scheme is the *brainVolume(interpretation, core) object*. These *objects* are simply regions of brain tissue that are defined in terms of a specific publication's parcellation scheme, providing the geometrical substrate for the data in the system. According to this treatment, brain atlases are considered to be "just another publication" with a parcellation scheme expressed as *brainVolume(interpretation, core) objects* to which we link other publications' *brainVolume* objects with relations (see later). Each subobject of any given *neuronPopulation(interpretation, core)* objects may be linked to a given *brainVolume(interpretation, core)* object in order to express its location.

This representation of neuroscientific data is very general and consequently may be used as a template for more specialized representations. We will illustrate different forms of *neuronPopulation* objects that are specific to individual domain-type/knowledge-type pairs in the next section. Additionally, many aspects of the structure of the NeuroScholar model are very similar to those encountered in the design of the USC Brain Project's federated scheme for repositories of experimental data (NeuroCore; see Part 3). In order to make the most of this similarity, we converted the data model of the NeuroCore federation of databases into our object-oriented system in order to make use of NeuroCore's widespread applicability and other desirable design features.

Building Rule Sets with Relations

Objects define separate things that may be compared. In the case of neuroscientific data, a good example of this derives from the process of converting the parcellation scheme of one paper to that of another. Stephan and Kötter mathematically formalize this process using set theory so that data can be computationally translated between different cortical maps (a procedure referred to as the Objective Relational Transform, or ORT; see Stephan *et al.,* 2000b). They use one of four rules to describe how the representation of a cortical area in

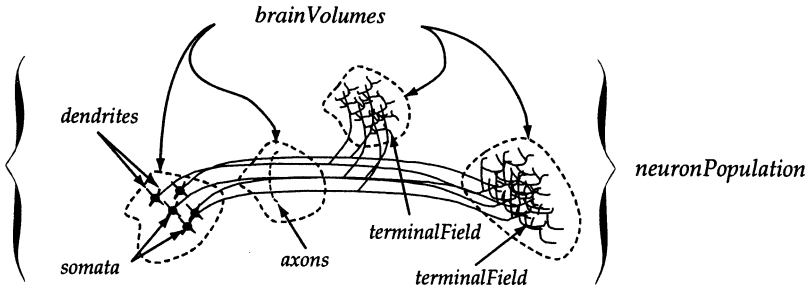

Figure 4 The general *neuronPopulation(interpretation, core)* object (see text for description).

one map relates to a similar representation in another map. Given two areas (A and B) in two maps (1 and 2), the rules would take the following forms:

1. "Area A in map 1 *is equivalent to* area B in map 2."
2. "Area A in map 1 *is enclosed by* area B in map 2."
3. "Area A in map 1 *encloses* area B in map 2."
4. "Area A in map 1 *overlaps* area B in map 2."

Stephan and Kötter develop this framework further by using an algorithmic approach to automatically translate data between maps or to define hybrid maps that use a mixture of nomenclature from different maps.

We use these spatial relationships in NeuroScholar to compare *brainVolume* objects from different parcellation schemes and to describe how data from a given paper may be related to the parcellation scheme in an atlas. Authors often describe a specific location in a paper with an alphanumeric abbreviation denoting the region of interest, and the process of linking data from different papers together is based on linking the regions of interest in the papers to standard brain structures defined in an atlas. For example, Allen and Cechetto performed a combined electrophysiological and tract-tracing study in 1993 where they stimulated sites in the lateral hypothalamus with injections of excitatory amino acids and then performed tract-tracing experiments by injecting tracer into stereotaxically equivalent regions in the contralateral hypothalamus (Allen and 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 (LHAt)" 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; see Fig. 5). We can capture this detail by generating relations and use these relations to translate any description of the properties of neurons within that region to a standardized parcellation scheme. It would be possible to allow users to trace the exact position of their interpretation of the *brainVolume* on the atlas through the use of a drawing tool for storage and subsequent retrieval (such as the NeuARt data browser; see Chapter 4.3).

In this case, these rules can be widened to accommodate other NeuroScholar objects as well. The equivalent relations of "*is equivalent to*," "*is enclosed by*," "*encloses,*" and "*overlaps*" when applied to pairs of *neuron Populations* are "*is equivalent to*," "*is a subset of*," "*is a superset of,*" and "*overlaps*." These relations could be used to link *neuronPopulation* objects in different parts of the knowledge base.

Making it Make Sense to Users: The Use of Annotations

Annotation primitives represent personalized human interactions between users and the contents of the system. They are designed to be read subsequently by other users of the system. There are three types of *annotation primitive,* and each one fulfills a different role. *Comments* may be attached to any *primitive* in order to express the user's opinion concerning that *primitive*. *Justifications* must be attached to *objects* that are based on a *user's interpretation* and simply explain the logic that supports the definition of the *object*. *Judgments* must be entered in order to support a chosen stance where a user selects one *relation* over another when contradictions occur.

6.3.4 Neuroscholar in Detail

In order to describe more detailed aspects of the system, it may be useful to describe how we would represent

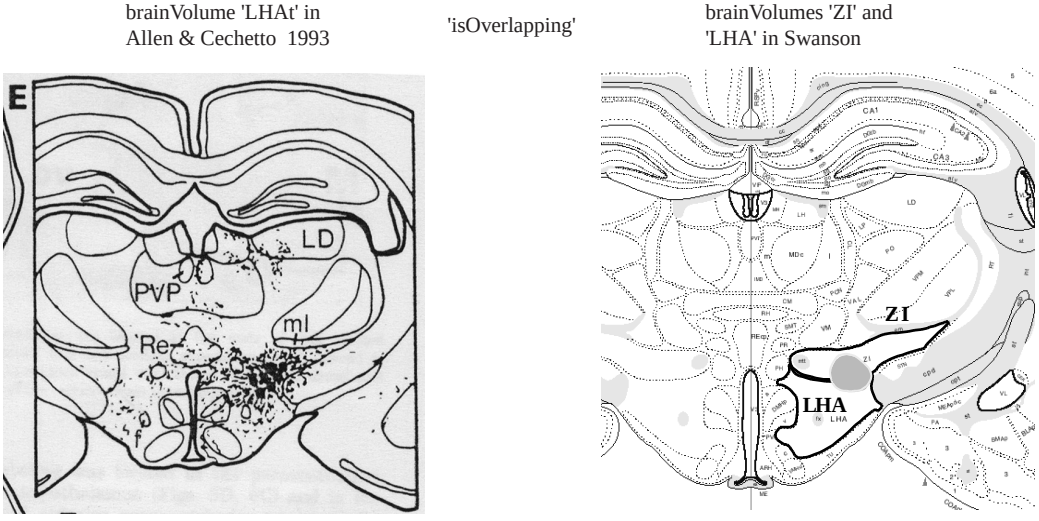

Figure 5 An example of a spatial relation.

data from a type of specific experiment in detail and then extrapolate to show how this approach could be generalized to other experiments. In this section, we discuss NeuroScholar's representation of tract-tracing experiments from a conceptual perspective.

A Short Introduction to Tract-Tracing Experiments

All tract-tracing experiments are based on the same basic design. An experimental animal is anesthetized 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 arbors. Then the tracer is 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 is sectioned and stained to reveal a pattern that shows the distribution of transported label (Blackstad *et al.,* 1981).

The end product of a given tract-tracing experiment is a three-dimensional map of tracer concentration at the time of the animal's death. Because tracer is originally applied in a localized injection to a small region of the brain, labeled regions of the brain must be connected to the injection site. The data are usually presented as photographs or camera lucida drawings of sections which show the distribution of label throughout the brain (for example, see Fig. 4 of Sesack *et al.,* 1989).

Studies are classified into "retrograde" and "anterograde," based on the type of axonal transport mechanism mediating the tracer's movement (Vallee and Bloom, 1991). Tracers that have been classified as "retrograde" are assumed to be taken up by axon terminals and then transported along the axons to the soma of neurons projecting to the injection site. In contrast, tracers that have been classified as "anterograde" are assumed to be taken up by the soma and dendrites of neurons and then transported to label the axonal terminals of that cell.

Unfortunately, interpreting labeling patterns is not as simple as described above. The categorization of tract-tracing techniques as "anterograde" and "retrograde" is a simplification. In reality, the uptake and transport processes of the cells can give rise to different patterns of labeling in different circumstances. The sensitivity of the tracer can be described as the ratio of the amount injected to the amount transported. This property determines how much tracer is used in a given injection and, therefore, the size of a given injection site. Tracer sensitivity and other tracer properties are largely determined by the uptake and transport mechanisms of tracer by dendrites, soma, axons, or terminals of cells at the injection site.

This often gives rise to "fibers of passage" labeling, where axons passing through the injection site have taken up tracer and transported it to terminals or soma. Some tracer chemicals can be transported transynaptically, so it cannot be assumed that neuroanatomical tracers only label cells that are directly connected to the injection site (Cliffer and Giesler, 1988; Dado *et al.,* 1990; Sawchenko and Swanson, 1981). If tracer can be taken up by the dendrites of cells that lie on the periphery of the injection site, it may be possible for anterograde label to be transported to the terminals of cells whose soma lie outside the injection site. Transport of some tracers can also occur in both anterograde and retrograde directions. In some cases, tracer could conceivably be taken up by axonal terminals, transported retrogradely, and then transported anterogradely down a collateral branch of the neuron (see Fig. 3 of Warr *et al.,* 1981). When attempting to minimize spurious results and uninterpretable labeling patterns, neuroanatomists cross-reference their studies so that both "anterograde" and "retrograde" methods are applied to the problem in order to corroborate the presence of connections between structures.

Readers of neuroanatomical papers have to trust the papers' descriptions of injection sites and the absence of transynaptic labeling or labeling from damaged and undamaged fibers of passage. Some methods are more prone to error than others and may be treated with some element of suspicion (Sawchenko and Swanson, 1981). If the injection site's boundaries do not coincide exactly with those of the nucleus or area being studied, the labeling produced in a single experiment can only be regarded as a subsample of the total population of cells involved. It would be necessary to take this subsampling factor into account when performing quantitative studies where the number of cells projecting to or from any one region might be counted. The opposite problem is evident when attempting to label the connections of a small structure. If the technique being used produces large injection sites, then the zone of active uptake will extend beyond the boundaries of the nucleus under study and produce false-positive labeling from neighboring regions.

Some techniques are more sensitive than others, labeling cells that others miss (Wan *et al.,* 1982). This may imply that the cells labeled in these experiments are usually an underestimation of the total population of cells with active uptake zones in the injection site. One would normally assume that this subsampling does not suffer from systematic errors, but it is possible that some classes of cells may be more strongly labeled than others; for example, cells with very extensive terminal arbors might tend to take up more retrograde tracer than neurons with small trees.

There are many different tract-tracing techniques available to the experimental neuroanatomist, but they do not, in general, produce results that conform simply to an ideal "retrograde" or "anterograde" model. The most commonly used methods include the autoradio-

graphic technique using tritiated amino acids, horseradish peroxidase, horseradish peroxidase conjugated to wheat germ agglutinin, the most common fluorescent dyes, and *Phaseolus vulgaris* leuco-agglutinin. These methods have been extensively studied and reviewed by Bolam (1992), Heimer and Robards (1981), and Heimer and Zaborszky (1989); the reader is referred to these standard neuroanatomical texts for more information.

We do not consider lesion studies or electrical stimulation studies to be tract-tracing methods, as neither method involves injection of an actively transported substance. The earliest techniques for understanding connections were based on introducing damage to the brain structures under study and then using the Fink-Heimer silver-staining procedure to label damaged axons and terminals and indicate the course of projections from the damaged area (see description in Blackstad *et al.,* 1981). However, the nature of the lesion can affect the data results. Aspiration and radiofrequency lesions suffer from the problem that fibers passing through the area where the lesion is delivered also degenerate. Cytochemical lesions avoid this problem by selectively damaging cell bodies and leaving fibers of passage unaffected. Electrical stimulation experiments involve placing a stimulating electrode in one area and a recording electrode in another and then measuring responses at the recording site after passing current through the stimulating electrode. This technique is used relatively rarely (but see Finnerty and Jefferys, 1993).

Representing the Logical Structure of Tract-Tracing Injections in NeuroScholar

Based on the characteristics of tracers described in the last section and a working knowledge of the tract-tracing literature, we devised an object scheme for experimental data that is schematically illustrated for an ideal anterograde tracer in Fig. 6. A *neuronPopulation(description, tract-tracing experiment)* object that is based on a *description* of data from a tract-tracing experiment is characterized by the following list of subobjects:

1. An *injectionSite(description, tract-tracing experiment)* subobject (this has a *brainVolume(description, core)* sub-subobject denoting the enclosing region that contains it)
2. An array of different *labeling(description, tract-tracing experiment)* subobjects, each of which could correspond to different types of labeling in different regions (so that each member of the array has a *brainVolume(description, core)* sub-subobject)
3. The *neuronPopulation(description, tract-tracing experiment)* object linked to an *experimentalProtocol (description, tract-tracing experiment)* object describing all aspects of how the experiment was performed

If a different protocol is used, then the same subobjects are used to represent the data (i.e., an *injectionSite* object, an array of *labeling* objects, and an *experimentalProtocol* object), but the meaning of this information

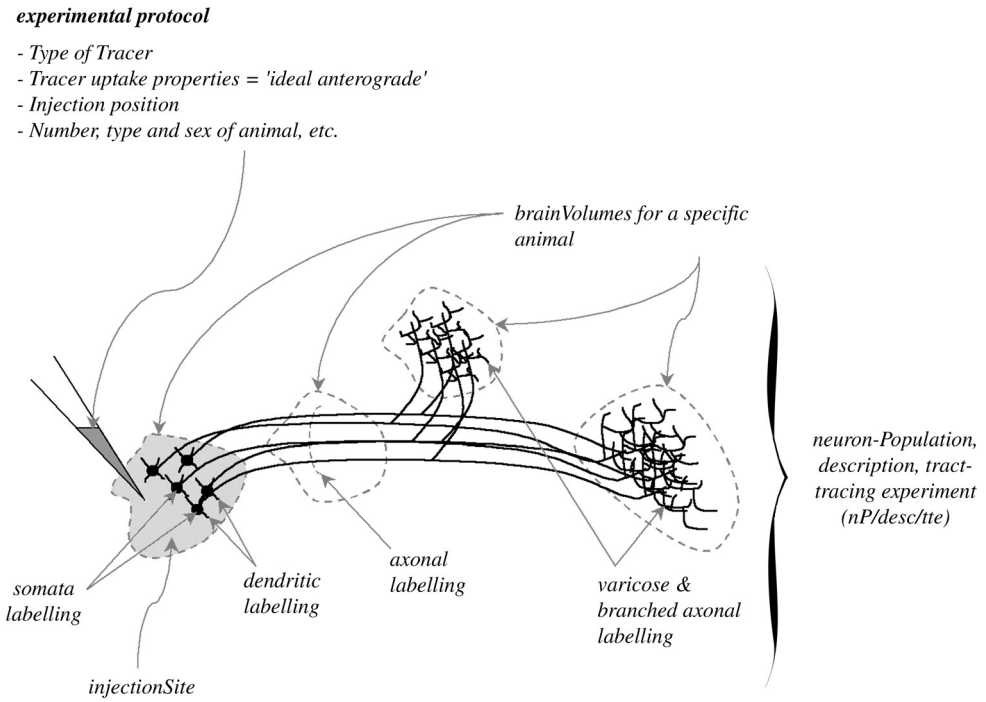

Figure 6 A *neuronPopulation(description, tract-tracing experiment)* object from an anterograde tract-tracing experiment.

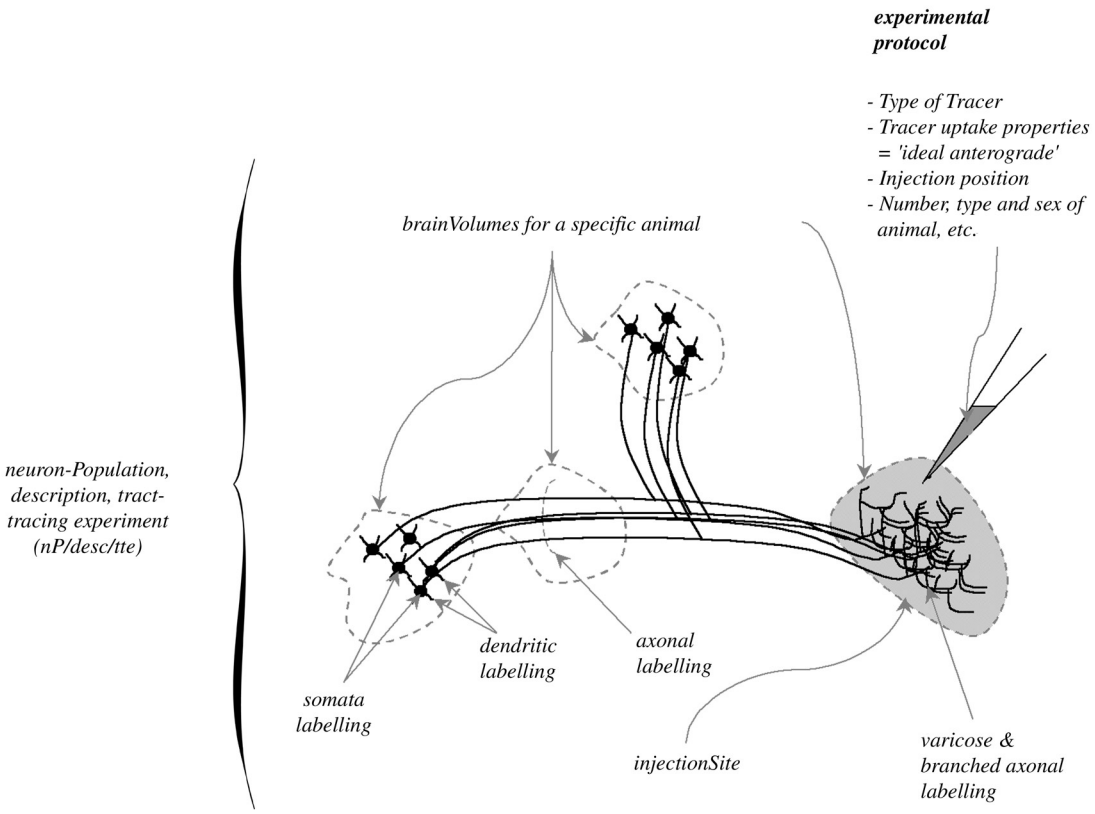

Figure 7 A *neuronPopulation(description)* object from a retrograde tract-tracing experiment.

may be completely different. A case illustrating the use of an ideal retrograde tracer is illustrated schematically in Fig. 7. Sadly, all retrograde tracers suffer from problems of uptake by fibers of passage (meaning that the tracer chemical is taken up by axons that are simply passing through the injection site without forming synapses), so a simple classification of tracer methods into ideal anterograde or retrograde types is unsatisfactory.

Instead, for each tracer chemical, we define an array of four values that denotes the strength of tracer uptake in the *dendrites*, *somata*, *axons,* and *terminalField* of a *neuronPopulation(interpretation, core)* object (see above). Thus, if we assume that PHAL tracers have ideal anterograde properties, we define the *tracerProperties* array of PHAL to be [*dendriteUptake* = "strong," *somataUptake* = "strong," *axonsUptake* = "none," *terminalFieldUptake* = "none"]. By the same token, we might describe the *tracerProperties* array of Fluoro-Gold to be [*dendriteUptake* = "none," *somataUptake* = "none," *axonsUptake* = "weak," *terminalFieldUptake* = "strong"]. At present, we define these arrays as designers of the system, but they could be derived from experimental data by deriving these arrays from NeuroScholar representations of the neuroanatomical experiments that define and test the methodology (e.g., Gerfen and Sawchenko, 1984; Wessendorf, 1991). This illustrates a very important aspect of the knowledge mechanics design strategy: to represent every logical stage of the reasoning process fully transparent and driven by experimental data (as much as possible). Practically, this may be defined within NeuroScholar by requiring that certain types of analysis require the results of other analyses as input data.

The results and conclusions of tract-tracing papers are most often concerned with "connections," the concept that one specified brain region contains neurons that project to and terminate in another region. This is a very commonly used concept, and we represent it in NeuroScholar by defining a new class of objects referred to as *blackBoxConnections*(*author interpretation, tract-tracing experiment*). These objects have three subobjects: two *brainVolume* subobjects denoting the regions of origin and termination of the connection and the *connectionStrength* object taking an ordinal value to denote the strength of the connection (i.e., "zero," "weak," "moderate," "strong"). A schematic illustration of this design is shown in Fig. 8.

The schematic descriptions we have laid out are simple illustrations of the data model being implemented. This is illustrated as an Entity-Relationship diagram in Fig. 9, which shows how the NeuroCore scheme was used a basis for the object-object relationships within the broader scheme of NeuroScholar's framework for knowledge mechanics (see Fig. 3).

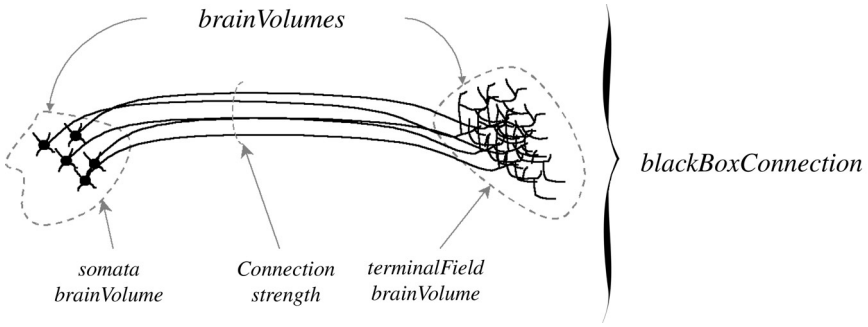

Figure 8 A *blackBoxConnection(author interpretation)* object from a retrograde tract-tracing experiment.

Broadly speaking, the object model may be broken into the following parts: (1) experimental meta-data (see the *expMetaData* objects in Fig. 9), made up of *interpretation* objects that capture a paper's knowledge; (2) research data (see the *researchData* objects in Fig. 9), which are made up of *description* objects from a paper's results section; and (3) *description* objects, which capture the experimental structure (see *experimentCore* objects in Fig. 9), the protocol and experimental manipulations (see *protocol* and *expManip* objects in Fig. 9), and details of the experimental subjects being used (see *researchGroup* and *researchSubject* objects in Fig. 9). There are two sets of "add-on" objects: *brainVolume(interpretation, core)* objects, which capture details of the parcellation scheme of the paper, and the *chemicalDictionary* objects, which capture details of the chemical properties of the tracer chemicals used.

Interpretive analysis procedures are represented in Fig. 9 as a dashed box superimposed on an arrow that leads from the children of the *researchData* object to the children of the *expMetaData* object. These procedures apply queries to the research data contained in a paper and then apply rules derived from the *tracerProperties* (see above) of the tracer chemical used in the experimen to generate *neuronPopulation(calculation, core)* and *blackBoxConnection(calculation, tract-tracing expt)* objects. These objects can be used to augment a user's interpretation of the data in a paper, especially where an author's own interpretations may be incorrect (if authors assumed that early tracer methods did not suffer from the fibers of passage problem, as was often the case).

More Relations in NeuroScholar

The so-called "strength" of neuroanatomical connections is often used to prioritize how different connections influence the global organization of a system (e.g., Scannell *et al.,* 1995), but the task of analyzing the interrelationships between descriptions of connections taken from different studies is difficult to approach from a formal perspective. Most neuroanatomical papers offer qualitative descriptions of connection strength rather than quantitative measurements. In the rat, quantitative data do exist for a limited number of connections that have been particularly well studied (Linden and Perry, 1983; Martin, 1986) but does not exist for the vast majority of connection reports. For example, the number of retinal ganglion cells that project to the superior colliculus is of the order 105 (Linden and Perry, 1983), and the weakest retinal efferents may only involve a few fibers, such as the retinal projection to the inferior colliculus (see Fig. 5 in Itaya and van Hoesen, 1982). In addition to this, the sensitivity of different neuroanatomical methods varies over at least two orders of magnitude (Ter Horst *et al.,* 1984; Trojanowski, 1983; Wan *et al.,* 1982). Furthermore, the density of label produced in any neuroanatomical experiment is dependent on the concentration and volume of tracer injected (Behzadi *et al.,* 1990).

A formal approach to generate finely graded connection strengths has been described that uses the basic premise that comparisons between labeling patterns representing connections within one experiment are very likely to be accurate, and comparisons between labeling patterns representing connections within one paper are quite likely to be accurate (Burns, *et al.,* 1996). Large numbers of these comparisons can be evaluated as a set of constraints for an optimization engine in order to calculate ordinal schemes of connections that best fit the data. Burns *et al.* used the MBOLTZMANN program to perform this process (Burns *et al.,* 1996; Hilgetag *et al.,* 1995). These comparisons can be built up within NeuroScholar as a set of ordinal relations ("is greater than," "is less than," etc.) between the *connectionStrength* subobjects of *blackBoxConnection* objects.

This process is a general model for a class of analyses, where the knowledge base can be queried to provide a large set of constraints between objects and then this set can be analyzed with optimization approaches such as MBOLTZMANN or optimal set analysis (OSA; see Hilgetag *et al.,* 2000) to give an accurate appraisal of the organization of the system within the bounds of the constraints defined by the literature.

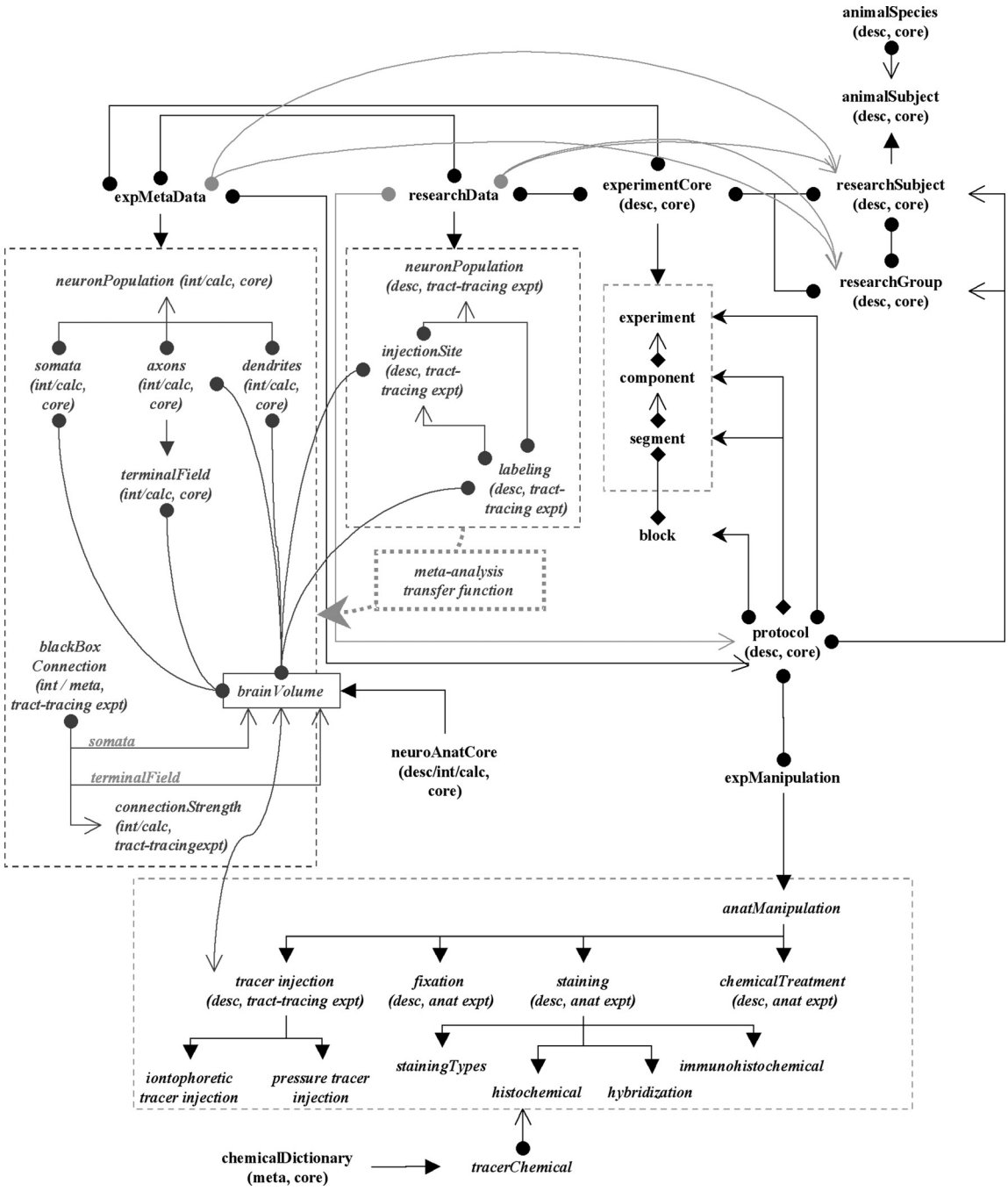

Figure 9 An entity-relationship diagram of the NeuroScholar representation of the NeuroCore database design. Arrows denote relationships where "——▶" represents a parent-child inheritance relationship and all other double-headed arrows denote an association. Multiplicity types of these associations are expressed as different types of arrowheads: Arrows ending in "——▶" represent a multiplicity of "0–1"; "——>" represent a multiplicity of "1"; "——◆" represent a multiplicity of "0–*n*;" and "——●" represent a multiplicity of "1–*n*."

Extending NeuroScholar to Other Data

NeuroScholar may be used as a knowledge mechanics framework for information from non-tract-tracing data from other neuroscientific experiments. Here, we describe electrophysiological or *in situ* hybridization data in the same illustrative framework as we used for tract-tracing experiments. As with tract-tracing studies, the key object in these studies will be the *neuronPopulation* object. In electrophysiological experiments, the *neuronPopulation* will be defined by the position of the recording electrode and will represent a "physiologically defined cell class," whereas in neurochemical experiments the *neuronPopulation* will be defined by the presence of labeling across the whole brain. These experiments may be broken into four parts:

1. *Preparation:* Training, lesions, conditioning, etc. (anything that an experimentalist might do to an animal before stimulating it within the experiment; see below)
2. *Stimulation:* A specific physiologically defined stimulus (such as an airpuff blown onto the cornea to elicit an eyeblink, an audible tone, or an injection of electrical current or a pharmacological agent)
3. *Cellular dynamics:* How a group of cells is functionally active electrophysiologically or neurochemically
4. *Behavior:* What the animal is doing

The causal relationships between these different parts of the system are shown in Fig. 10. These relationships have no significant relationship concerning the way that the information is handled in the database itself but may strongly influence the way that the information is handled within the system. If the four parts are related according to the flow diagram in Fig. 10, then several types of experiment may be defined according to Table 1.

Linkages between *neuronPopulations* from different experiments must be built on the basis of the actual criteria that exist within the literature. If the constituent *brainVolumes* of each of the *neuronPopulations* are located in the same region, then spatial relations can be used to link the data. If researchers have explicitly described linked phenomena within the paper (such as with double-labeling studies or by combining electrophysiological recording with cell-filling, for example) then the *neuronPopulation* objects themselves may be linked using set-theoretical relations.

6.3.5 The Significance of Knowledge Mechanics

The first issue of the journal *Trends in NeuroSciences* began with the phrase, "Even the most active neuroscientist spends more working hours on reading, reviewing, and writing scientific reports that on direct experimental effort" (Bloom, 1978). Eighteen years later, Floyd Bloom, the author of that quote, wrote in the same journal how he anticipated the need for knowledge-management systems concerned with making sense of published neuroscientific data (Bloom, 1995). Bloom and Young devised one of the earliest effective informatics solutions to the problem of representing as much of the breadth of neuroscientific data as possible, and their database (called Brain Browser) still provides a valuable reference work today (Bloom *et al.,* 1989).

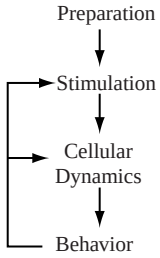

Figure 10 Flow diagram representing the causal interactions between parts of a physiological experiment in NeuroScholar.

Bloom's foresight in both recognizing the scope of the problem of applying informatics solutions to analyzing the literature and his conceptualization of the software required to tackle the problem cannot be faulted. In fact, many of the design features of NeuroScholar attempt to replicate some of the functionality of Brain Browser; however, we view the definition of NeuroScholar as an implementation of a theoretical approach called "knowledge mechanics" that is new and innovative. As this chapter has described, this approach employs simple, extendable data models made up of a relatively small number of data structures (so-called *fragments*, *objects*, *properties*, *relations*, and *annotations*) according to tested

Table 1

Code	Experiment Type	Description
C	Cellular Dynamics	Passively recording activity from an animal
SC	Stimulation–cellular dynamics	Recording from an animal under a specific stimulus paradigm
PSC	Preparation—stimulation—cellular dynamics	Recording from an animal under a specific stimulus when the animal has been "primed" in some way
PSCB	Preparation—stimulation—cellular dynamics—behavior	Recording from a behaving animal under a specific stimulus when the animal has been "primed" in some way
PCB	Preparation—cellular dynamics—behavior	Recording from a behaving animal when the animal has been "primed" in some way
PSB	Preparation—stimulation—behavior	Observing an animal"s behavior under a specific stimulus when the animal has been primed (e.g., lesion experiments)
B	Behavior	Simple observation of animal's behavior

software engineering principles to define a pragmatic ontology. Given that the user base of this software consists of scientists who have an inherent interest in understanding how their data, information, and knowledge are handled, we present the underlying design of this approach as a theoretical representation for neuroscientific knowledge so that the complex details of our computational representation are explicitly stated.

Finally, we believe that this approach can only be effective if it is incorporated into the process of publication itself and is used by a large number of scientists. Thus, the significance of the approach will depend heavily on its usefulness and predictive power as a theoretical approach and this will be affected by design features of the system's user interface and the quality of technology supporting the software. For appraisal of this and other aspects of our work, see the NeuroScholar Website at http://neuroscholar.usc.edu/.

References

Allen, G. V., and 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(3),** 421–438.

Behzadi, G., Kalen, P., Parvopassu, F., and Wiklund, L. (1990). Afferents to the median raphe nucleus of the rat, retrograde cholera toxin and wheat germ conjugated horseradish peroxidase tracing, and selective D-[3H]aspartate labelling of possible excitatory amino acid inputs. *Neuroscience* **37(1),** 77–100.

Blackstad, T. W., Heimer, L., and Mugaini, E. (1981). General approaches and laboratory procedures. In *Neuroanatomical Tract Tracing Techniques* (L. Heimer and M. Robads, Eds.). Plenum Press, New York.

Bloom, F. (1995). Neuroscience-knowledge management, slow change so far. *Trends Neurosci.* **18(2),** 48–49.

Bloom, F. (1978). New solutions for science communication problems needed now. *Trends Neurosci.* **1(1),** 1.

Bloom, F., Young, W., and Kim, Y. (1989). *Brain Browser.* Academic Press, New York.

Blum, B. (1986). *Clinical Information Systems.* Springer, New York.

Bolam, J. (1992). *Experimental Neuroanatomy,* A Practical Approach. Oxford University Press, New York.

Burns, G. A. P. C. (1997). Neural connectivity in the rat, theory, methods and applications. In *Physiology*, Oxford University Press, New York, p. 481.

Burns, G. A. P. C., and Young, M. P. (2000). Analysis of the connectional organisation of neural systems associated with the hippocampus in rats. *Phil. Trans. Roy. Soc. London B,* Biol. Sci. **255(1393),** 55–70.

Burns, G. A. P. C., O'Neill, M. A. and Young, M. P. (1996). Calculating finely-graded ordinal weights for neural connections from neuroanatomical data from different anatomical studies. In *Computational Neuroscience.* Trends in Research, J. Bower, Boston, MA.

Cliffer, K. D., and Giesler, G. J., Jr. (1988). PHA-L can be transported anterogradely through fibers of passage. *Brain Res.* **458(1),** 185–191.

Dado, R. J., Burstein, R., Cliffer, K. D., and Giesler, G. J., Jr. (1990). Evidence that Fluoro-Gold can be transported avidly through fibers of passage. *Brain Res.* **533(2),** 329–333.

Felleman, D. J., and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1(1),** 1–47.

Finnerty, G. T., and Jefferys, J. G. (1993). Functional connectivity from CA3 to the ipsilateral and contralateral CA1 in the rat dorsal hippocampus. *Neuroscience* **56(1),** 101–108.

Gerfen, C. R., and 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* leukoagglutinin (PHA-L). *Brain Res.* **290,** 219–238.

Heimer, L., and Robards, M. J., Eds. (1981). *Neuroanatomical Tract-Tracing Techniques.* Plenum Press, New York.

Heimer, L., and Zaborszky, L. Eds. (1989). *Neuroanatomical Tract-Tracing Methods.* Vol. 2. *Recent Progress.* Plenum Press, New York.

Hilgetag, C., Burns, G., O'Neill, M., Scannell, J., and Young, M. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Phil. Trans. R. Soc. London B* **335(1393),** 92–110.

Hilgetag, C. C., O'Neil, M. A., Scannell, J. W., and Young, M. P. (1995). A novel network classifier and its application, optimal hierarchical orderings of the cat visual system from anatomical data. In *Genetic Algorithms in Engineering Systems,* Innovations and Applications.

Itaya, S. K., and van Hoesen, G. W. (1982). Retinal innervation of the inferior colliculus in rat and monkey. *Brain Res.* **233(1),** 45–52.

Linden, R., and Perry, V. H. (1983). Massive retinotectal projection in rats. *Brain Res.* **272(1),** 145–149.

Martin, P. R. (1986). The projection of different retinal ganglion cell classes to the dorsal lateral geniculate nucleus in the hooded rat. *Exp. Brain Res.* **62(1),** 77–88.

Merriam Webster (2000). *WWWebster Dictionary* http://www.m-w.com/netdict.htm.

Muller, R. J. (1999). *Database Design for Smarties,* Using UML for Data Modeling. Morgan Kaufmann, San Francisco, CA.

Pressman, R. (1992). *Software Engineering, A Practitioner's Approach.* McGraw-Hill, New York.

Rational Software Corp. (1997). *UML Semantics* Version 1.1. Rational Software Corp., Santa Clara, CA, www.rational.com.

Sawchenko, P. E., and Swanson, L. W. (1981). A method for tracing biochemically defined pathways in the central nervous system using combined fluorescence retrograde transport and immunohistochemical techniques. *Brain Res.* **210(1–2),** 31–51.

Scannell, J. W., Burns, G. A. P. C., Hilgetag, C. C., O'Neil, M. A., and Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cerebral Cortex* **9,** 277–299.

Scannell, J. W., Sengpiel, F., Benson, P. J., Tovée, M. J., Blakemore, C., and Young, M. P. (1996). Visual motion processing in anterior ectosylvian sulcus of the cat. *J. Neurophysiol* **76(2),** 895–907.

Scannell, J. W., Blakemore, C., and Young, M. P. (1995). Analysis of connectivity in the cat cerebral cortex. *J. Neurosci.* **15(2),** 1463–1483.

Sesack, S. R., Deutch, A. Y., Roth, R. H., and Bunney, B. S. (1989). Topographical organization of the efferent projections of the medial prefrontal cortex in the rat, an anterograde tract-tracing study with *Phaseolus vulgaris* leuco-agglutinin. *J. Comp. Neurol.* **290(2),** 213–242.

Stephan, K. E., Hilgetag,, C. C., Burns, G. A. P. C., O'Neill, M. A., Young, M. P. and Kötter, R. (2000a). Computational analysis of functional connectivity between areas of primate cerebral cortex. *Phil. Trans. R. Soc. London B* **355(1393),** 111–126.

Stephan, K. E., Zilles, K., and Kötter, R. (2000b). Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Phil. Trans. R. Soc. London B* **335(2393),** 37–54.

Swanson, L. W. (1998). *Brain Maps, Structure of the Rat Brain.* Elsvier Science, Amsterdam.

Ter Horst, G. J., Groenewegen, H. J., Karst, H. and Luiten, P. G. M. (1984). *Phaseolus vulgaris* leuco-agglutinin immunohistochemistry, a comparison between autoradiographic and lectin tracing of neuronal efferents. *Brain Res.* **307(1–2),** 379–383.

Trojanowski, J. Q. (1983). Native and derivatized lectins for *in vivo* studies of neuronal connectivity and neuronal cell biology. *J. Neurosci. Methods* **9**.

Vallee, R. B., and Bloom, G. S. (1991). Mechanisms of fast and slow axonal transport. *Annu. Rev. Neurosci.* **14,** 59–92.

van Heijst, G., Schrieber, A., and Wielinga, B. (1997). Using Explicit Ontologies in KBS development. *Int. J. Human-Computer Studies* **45,** 183–292.

Wan, X. C. S., Trojanowski, J. Q., and Gonatas, J. O., (1982). The dynamics of uptake, transport and clearance of horseradish-peroxidase (Hrp). conjugates of cholera toxin (CtHrp). and wheat-germ agglutinin (WgHrp). *J. Neuropathol. Exp. Neurol.* **41(3),** 350.

Warr, A., Olmos, J. D., and Heimer, L. (1981). Horseradish peroxidase, the basic procedure. In *Neuroanatomical Tract Tracing Techniques* (Heimer, L., and Robads, M., Eds.). Plenum Press, New York, pp. 207–262.

Wessendorf, M. W. (1991). Fluoro-Gold, composition, and mechanism of uptake. *Brain Res.* **553(1),** 135–148.

Young, M. P. (1993). The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. London B,* Biol Sci. **252(1333),** 13–8.

Young, M. P. (1992). Objective analysis of the topological organization of the primate cortical visual system [see comments]. *Nature* **358(6382),** 152–155.