Foundations of a Structured Approach to Characterising Domain Knowledge

R. C. Paton, H. S. Nwana, M. J. R. Shave,
T. J. M. Bench-Capon & S. Hughes.

The MEKAS Project
Department of Computer Science
University of Liverpool
Liverpool, L69 3BX
U.K.
email: rcp@cs.liv.ac.uk

Abstract

One of the key phases in Knowledge Based Systems (KBS) construction is Knowledge Acquisition. However, human knowledge about domains is so complex that without an *analysis* stage that probes the underlying nature of the real world problem and how human experts conceptualise it, the knowledge incorporated within a KBS remains shallow and incomplete. In this paper, we highlight foundational details of a structured approach to knowledge analysis and describe its application to domains associated with software installation and neural networks.

1 Introduction: Domain Characterisation and Knowledge Acquisition

In this paper we present the foundations of an approach to the characterisation of problem solving domains for the development of knowledge based systems. These foundations come from a broadly based, multi-disciplinary perspective which includes the cognitive sciences, computer science and the philosophy of science. The breadth of disciplines required reflects the need to deal with the complexity of knowledge that needs to be acquired when carrying out knowledge acquisition.

Knowledge acquisition should be a key stage in any methodology which seeks to design and construct a non-trivial knowledge based system (KBS). However, it represents a major hurdle in building such systems. It usually involves eliciting, analysing and interpreting the language that the expert uses when solving a problem and then transforming it into a suitable machine language. As such, knowledge acquisition is crucial since the power and utility of the resulting knowledge based system depends on the quality of the underlying representation of expert knowledge (Kidd, 1987). Furthermore, as a recent comprehensive United Kingdom survey reported, KBS developers did not consider knowledge elicitation from the expert a problem in itself, rather, it was making sense of the great mass of detail and information obtained, so that it can be organised and represented (O'Neill & Morris, 1989). We refer to this process of organising knowledge (i.e. the raw data) gained from human experts, literature, manuals, journals and other sources into a coherent and unambiguous structure of the domain as knowledge analysis; we further agree with O'Neill & Morris that it is this process which can facilitate a better understanding of the requirements of knowledge acquisition. It is pertinent to note at this point that the process of knowledge acquisition is complex and fraught with difficulties. Some of these are addressed in this paper and certain solutions are suggested.

In order to clarify some of the reasons for problems in knowledge acquisition, it will be valuable to look briefly at the history of the subject. In general terms knowledge acqusition has focussed primarily on the transfer of problem solving expertise to computer based software called a knowledge based system. First generation approaches to the development of knowledge based systems are mainly of two types: ‘Stage Based’ and the prototyping approach. Stage based approaches present life cycle definitions based on ideas from conventional software engineering. For example, Buchanan *et al.* (1983) proposed a life cycle definition including the following stages: identification of the problem to be investigated, conceptualisation of a model to represent the knowledge, formalisation, implementation of the computer based software and testing/revision. An example of the prototyping approach is that of Guida & Tasso (1989). They propose a plausibility study, rapidly followed by a demonstration prototype, development of the full prototype, development of target system, operation and maintenance/revision. These approaches have similarities to conventional data analysis methodologies of computing such as Structured Systems Analysis and Design Methodology or SSADM (Downs *et al*., 1986). Clearly, these methodologies are very general as they attempt to address the entire knowledge engineering endeavour. In so doing, they fail to acknowledge knowledge acquisition as a separate phase of enquiry. Indeed, prototyping, which has been much used to date, may not fully utilise the knowledge acquisition process because the underlying assumption is that one can uncover the structure of the expertise of the domain at a very early stage with little or no analysis. A key criticism of these approaches is that they do not acknowledge the need to analyse the domain prior to the development of the KBS.

A second generation of approaches to KBS development are emerging which address the issue of analysis in terms of knowledge-level analysis (Newell, 1982). Knowledge-level analysis refers to the modelling of an intelligent system's problem-solving behaviour independent of whatever symbols might ultimately be used to program these behaviours in a computer. These approaches include: Ontological Analysis (Alexander *et al.*, 1986), the approach through Generic Tasks (e.g., Clancey, 1986 and Chandrasekeran, 1985) and the KADS (Knowledge Acquisition, Documentation and Structuring) methodology (Wielinga *et al.*, 1988). For example KADS, which is probably the most established modelling approach in Europe, seeks to develop a computer-based model of problem solving behaviour for a particular set of tasks. The authors suggest a four layer architecture which describes the domain, the type of inferences, the tasks, and strategic structures. They also describe a number of domain-independent conceptual primitives used for representing the inference and task layers. These primitives, called interpretation models in KADS, are similar to Clancey's (1986) generic tasks which depict the levels where knowledge-level analysis is carried out (eg heuristic classification). The common feature of these approaches is that they are concerned with modelling at the knowledge level. As such they do not provide a model of human competence at a task nor do they develop a model which relates the human problem solving behaviour or human communication about the domain in the real world.

We argue that a fuller appreciation of the nature of a domain and its relationship to the real world can assist a knowledge engineer in knowledge acquisition. The descriptions of domains which are required, must be grounded or contextualised in human social and cognitive processes (Gaines, 1989) as well as in the real world (Clancey, 1989). We have suggested that the *characterisation* of a domain, which provides its context and reference to the real world, is the central goal of knowledge analysis (Paton & Nwana, 1990a). In saying this we wish to demonstrate that domain characterisation is an important goal in the early stages of a KBS development project. The need for this kind of analysis has emerged from a variety of industrially-sponsored knowledge acquisition projects based at the University of Liverpool (e.g. Finch, 1989; Hughes, 1986; Watson *et al.*, 1989) and is further supported by the requirements of two major industrial collaborators.

The issues that have been identified from these investigations, together with many others in this area of research, show why knowledge acquisition remains difficult. In particular, key issues associated with characterisation have not been addressed. These include:

- The lack of any deep theory of knowledge acquisition that has an explicit statement about its philosophical underpinning (Bradshaw & Woodward, 1989);
- The need to provide an adequate cognitive definition and understanding of the domain (Shaw, 1989);
- A means by which an expert and knowledge engineer can harmonise their mental models (Recogzei & Plantinga, 1987); especially as they seek to communicate with each other and talk about the same things in the real world.
- The need for guiding principles which allow a knowledge engineer to navigate a domain and make sense of the mass of information obtained (O'Neill & Morris, 1989);
- A way to match elicitation techniques to the nature of a domain. Despite numerous tools (see Boose, 1989), a major obstacle is that little guidance is available in this area (Kitto & Boose, 1989).
- The need for knowledge acquisition in its early stages to focus on the domain, rather than be driven by tools which are in turn driven by implementation concerns (Woodward, 1989; Paton & Nwana, 1990a).

Current approaches to KBS development often move too quickly to the stages of design and implementation of a computable artefact without first providing an adequate description (analysis) of the nature of the domain. In an attempt to provide answers to the problem of acquisition, we have proposed, and are developing, a methodology for knowledge analysis which produces detailed characterisations of the nature of domains.

By ‘domain’, we refer to that body of knowledge which bears on the underlying problem, present and future, which is identified in relation to the task, and which enables that task to be brought to solution (see Shapere, 1977). It is argued that domains of knowledge are the products of human abstraction and that, as with any abstraction-product, their formation depends upon a theory. For the purposes of domain characterisation, we have proposed in Paton & Nwana (1990) that there are seven fundamental characteristics of a domain:

- Theory - the conceptual framework used to construct and maintain a domain.
- Metaphor - language used to describe the domain, especially in global terms.
- Metatheoretical constraints - fundamental concepts such as time and causality.
- Relations with other domains - similar bodies of knowledge.
- History - the evolution of the domain in time.
- Structure - the parts, relations and organisation of the domain.
- Purpose - the problems which the domain addresses, in terms of their solution.

A fuller justification for these seven features and their interrelationships can be found in Paton *et al* (1991c). In order to provide a focus for discussion, this paper is limited to particular aspects of the theoretical and metaphorical nature of domains. We present foundational details of our approach and from this develop certain techniques for analysing domain knowledge. These techniques are an integral part of a structured approach to knowledge analysis which is under development.

2 Humans Build Models in order to Understand the Real World

> " Men think in terms of models." (Deutsch, 1951)

This is the fundamental assumption of our approach. In order to manage what we know about the real world we must simplify its complexity. We achieve this by abstraction and we call the products of this process *models*. In saying this, it is important to re-emphasise that a domain is that body of knowledge required when solving problems. The cognitive products generated to bring about particular solutions are models.[1]

We approach knowledge analysis with the observation that humans structure their knowledge and that an understanding of how structuring processes take place, what is produced and how changes take place over time, will provide clues to the way domains are organised. Essentially, we agree with Forrester (based on Bruner); that a theory:

---

[1] The term model is used because of its consistency with the term used by philosophers of science (e.g., Hesse, 1963; Harré, 1970; Rothbart, 1984; Aronson, 1991) and by the computing community (e.g., Lehman, 1977; Lewis & Smith, 1979, Kangassalo, H., Ohsuga, S. & Jaakkola, H., 1990). It is also consistent with the approach to metaphor which we elaborate on below, see also, Soskice, 1985; Rothbart, 1984; Way,1991). As such, model is preferred to, for example, ‘belief system’ as the latter carries implicit ideas about the nature of knowledge which we do not address in this paper (e.g., the notion of ‘system’ in the context of belief system).

> "... is essential if we are to effectively interrelate and interpret our observations in any field of knowledge. Without an integrating structure, information remains a hodge podge of fragments. Without an organising structure, knowledge is a mere collection of observations, practices and conflicting incidents." (Forrester, 1968, page 1-2).

Hence, the key feature of our approach to the nature of domains as conceptualised by human experts is that they are integrated structures. In saying this we argue that knowledge is organised into frameworks. This has a profound implication for our approach in that we can anticipate possible conceptualisations which people may have of a domains (i.e. through knowledge of how they are constructed and maintained) before we embark upon knowledge acquisition.

We use theory to refer to both a formal body of knowledge recognised by a community of people (as in the scientific sense) and to any framework of beliefs which is integrated in some way so as to provide a person with solutions to problems. In saying this, we differentiate model from theory in that the latter is more general than the former and that a model is the cognitive product of a theory. Theories are essential to the kinds of ontologies we possess and to the ways we represent or model the world when solving problems (Carey, 1985; Karmilloff-Smith, 1988; Keil, 1989). Furthermore, as Medin & Wattenmaker (1987) suggest, conceptual coherence is theory-based and is derived from both the internal structure of the conceptual domain and the position of the concept in the complete knowledge base. This is extremely important to our view of how knowledge is organised. What we believe to exist (ie our ontology) relates to a theory which defines it. The importance of this relationship, between theory and ontology, is not only supported by realist philosophers of science (eg Aronson, 1984; Bhaskar, 1978) but also by developmental psychologists (eg Carey, 1985). The metaphorical nature of knowledge also provides details of theoretical entities as we seek to specify the nature of a domain (see Black, 1979) so that an abstraction of a real world complexity is partly constructed by relating what there is to a set of metaphors which are deeply embedded in our thought and language (Pepper, 1928; Lakoff & Johnson, 1980). Thus, our conceptualisation of a model shows that it is theory-dependent and an understanding of its theory-dependency can help us anticipate the structure of a domain. The emphasis of the approach to knowledge analysis described in this paper is concerned with the theories people use to construct, maintain and change problem solving domains. In order to appreciate certain properties of theories in our approach, we have imported relevant ideas from the history and philosophy of science.

3 Towards a Characterisation of Theories

Theories may be thought of as frameworks of concepts tied together by propositions or sets of propositions. The purpose of such frameworks is to provide the means for description, prediction, explanation, identification, classification and diagnosis. They act as the basis for hypothesis generation and form the intellectual and socially acceptable structures by which the scientific enterprise continues. We now consider some particular features of scientific theories from which we are developing analytical techniques for domain characterisation.

3.1 Scientific Realism as a Basis for the Characterisation of Theories

We adopt a realist perspective for understanding the nature of domains. In simple terms, we believe that there is more to the world than what is perceived by the senses and formed into a theory, and more than what is present in the mind. In saying this, we do not seek to refute the empiricist or idealist positions nor debate the adequacy of realism in science; these are issues for philosophers. We come with a basic assumption, the world exists independently of our senses and mind and we use theories to inform us about its existence and nature (Aronson, 1984).

It is now possible to describe (model) the production of models. In the simplest case a model is produced by abstracting from what is observable alone (left-hand side of Figure 1). However, this kind of model will lack explanatory power because it cannot account for the causal relations between its parts. The only explanations available can be no more than variants of statistical associations. Most scientific models have explanatory power. This is because the descriptive or homeomorphic model (see Figure 1) is not only formed by abstraction from observables but is also dependent on models of what cannot be observed.

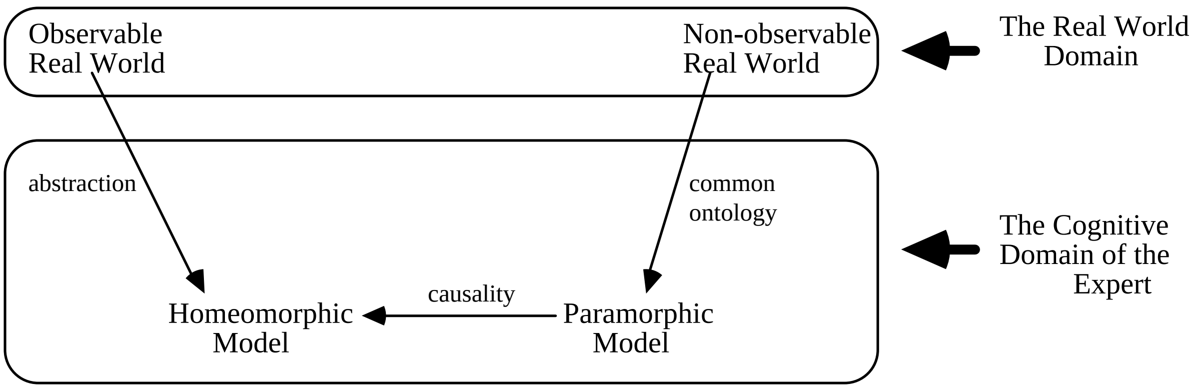


Figure 1 - A Summary Scheme of Model Relations (adapted from Harré, 1970)

These explanatory (paramorphic) models provide the causal framework necessary for explanation. Domains which exhibit explanatory capabilities in this way will have relational or iconic properties (due to their causal infrastructure). If an expert can provide explanations in this way, we can anticipate a theory which will have iconic (relational) properties (Harré, 1986). This is important when the dynamic aspects of the domain come to be modelled. The basis for paramorphic (explanatory) models is in interpretations of the unobservable real world which share common kinds of entities. It is metaphors which provide the context for such common ontologies (see Aronson, 1984), that is, common kinds of entities, and this is often noted when ideas are imported from related domains (Paton *et al.*, 1991a). Our adherence to this perspective provides us with a framework for making certain expectations about the ontological nature of domains concerned with time, causality, categorisation and mechanism (see Bhaskar, 1978).

3.2 Properties of Theories

Theories can be typified according to certain basic properties and, following the work of Harré (1986), we identify three basic types of theory:

- **Type 1 theories** are cognitive objects with pragmatic properties that enable the constitution, classification and prediction of observable phenomena. A typical Type 1 theory is Newtonian mechanics. [Note: these theories lack explanatory power and are similar to heuristics]. They can be characterised by the left-hand side of Figure 1. Thus, in the social sciences ethnomethodology yields Type 1 theories of social interaction by looking for patterns among observable actions, including verbal actions. The ontologies of Type 1 are, therefore, only observables.
- **Type 2 theories** are cognitive objects with iconic properties that enable the representation (including sometimes the picturing) of a certain class of unobservable beings. Typical Type 2 theories are the bacterial theory of disease, plate tectonics and X-ray stars. These theories involve the representation of a physical system and its modes of behaviour, which, at the time of the formulation of the theory, have not yet been observed. The vast majority of scientific theories are of Type 2. These theories make use of common ontologies whose context is often supplied by key metaphors. For example, the realist notion of a virus as a disease-causing particle was postulated in the mid-nineteenth century but was only confirmed to exist in the mid-twentieth century with the invention of electron microscopes. Its basis in theory existed prior to its confirmation with scientific instruments. The ontologies of Type 2 are both observable and non-observable.
- **Type 3 theories** are cognitive objects with mathematical properties which enable representations of non-picturable systems of beings and of their behaviour and interrelations. The ontology behind abstractions such as symmetry, transformation and harmony is a case in point, and we note the importance of what could be called the relationship between ontology and "mathematical beauty", from the Pythagoreans to the present time (see Engler, 1990). Einstein's theory of special relativity is an example of this type. The ontologies of Type 3 are both observable and non-observable but the non-observable referents may never become observable.

Hence, from our point of view, theories have certain fundamental properties which we may identify as *pragmatic* and/or *iconic* and/or *mathematical*. This is not to say that over time a theory may change so that, for example, the Type 3 nature may become Type 2. For example, the initial idea of a virus was Type 3, a disease-causing agent, which was later described in Type 2 terms as a disease-causing particle. (Note: the difference between agent and particle is non-trivial).

Kidd (1987) proposed a classification of domains which bears a resemblance to Harré's categorisation although we note that Kidd emphasises the need for a language, i.e. a knowledge representational medium. Harré focuses on theory which not only gives a means of representation but also provides conceptual techniques for analysis. Harré's categorisation includes the following implications for knowledge analysis:

- Sometimes experts will be able to provide no more than type 1 theories, i.e. rules of thumb or heuristics. The knowledge engineer can anticipate this: it is naive to think experts can explain all their knowledge. It was indeed the case we found in an investigation we carried out in a domain of neural networks. For example, the number of units or hidden layers in a neural network had no explanatory basis in any theory. This may seem a simple observation but it is worth noting that it is clearly an implication of type 1 theories.
- If the theory the expert articulates is more of a type 2 theory then iconic properties can be anticipated and the explanatory models used investigated. It is in this vein that we have proposed and are using metaphorical graphs and imagery to gain further insights into the nature of the theory and hence domain. Metaphorical graphs are simple paper representations of the systemic structure of a domain (see Figure 3).
- Type 3 theories are rarer in most real domains we seek to analyse. As such domains are non-picturable it is more difficult to gain insights into their nature. However, certain important ideas, which appear in experts language, are relevant to ontologies associated with Type 3 theories and include, symmetry, harmony, order, coherence, unity, elegance and simplicity. Note that these reference some common assumptions people have about the world. As we have seen, this is manifested in the ontologies people use.

It is important to note that theories change over time and, their characterisation according to types is not always clear-cut. However, the value of this classification to a knowledge engineer is that it can be used to evaluate models of a domain and to suggest the kinds of knowledge that are required for further acquisition. In all they provide a valuable way of helping to organise and document the large amount of information acquired from an expert.

3.4 The Importance of Metaphor to Theory Development

The metaphorical context in which analogies and inter-theoretical conceptualisations are made is an important aspect of the way domains are organised. Boyd (1979) and Kuhn (1979) highlight a key role metaphor plays in establishing links between the language of science and the world it purports to describe. Morgan (1980) notes that scientific theories are constructed as symbolic forms and the process of metaphorical conception is a basic mode of symbolism. Thus scientists view the world metaphorically as they go about their research. Metaphors help make up how we see the world, how we set about studying it and how it can be understood in terms of ways of perceiving relationships and situations from different perspectives (e.g. Genter & Grudin, 1985). As such, they are fundamentally important to the ways people go about constructing domains. Some metaphors are related to a domain as a whole and provide global details, others are related to specifics.

We have identified some basic kinds of global metaphor which are important in knowledge analysis; two in particular are systemic metaphors and spatial metaphors (see also Paton *et al.*, 1990). Specifically, the former provides information about the parts, inter-relations and organisation of a system and the latter helps us understand its functionality. For example, a neural network (in the computing sense) may be thought of as a machine which transforms inputs into outputs through some mechanism although its behaviour may be described in terms of changes to a landscape. The two basic kinds of metaphor are abstract in nature and we may identify further metaphors associated with each. A network showing some of the top-level features of our categorisation of global metaphors is shown in Figure 2. A concept in **bold** refers to a global metaphor, and the associated concepts which are listed are properties of the metaphor (we call them M-properties).

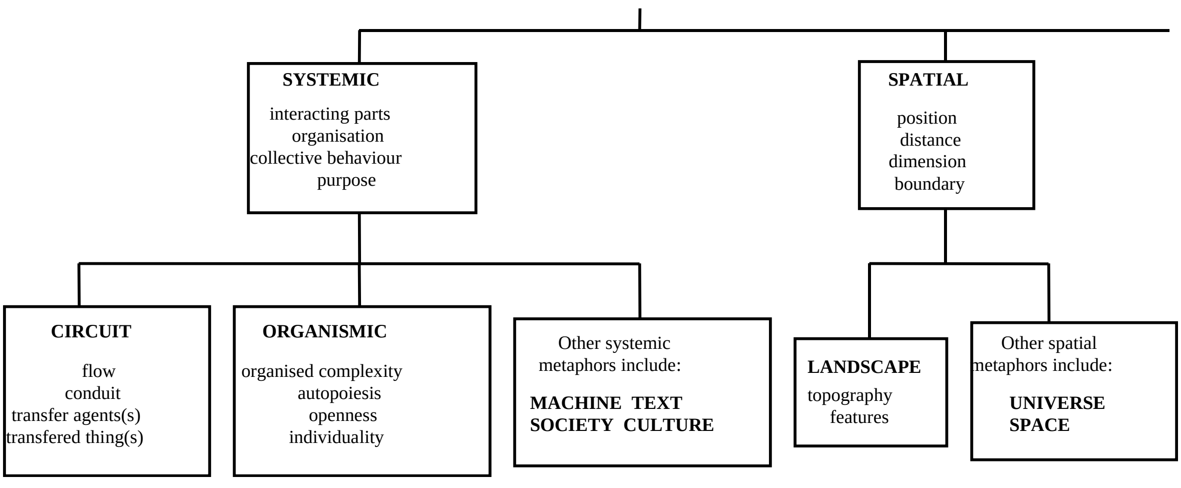


Figure 2 - Some Global Metaphors

It was noted in section 3.2 that type 2 theories have iconic properties. One application of this relates to iconic representations of systemic metaphors which we call 'metaphorical graphs'. Some simple forms of these are shown in Figure 3 to indicate some of their iconic and metaphorical relations, and an example of their use is reported in section 4.2.

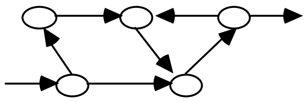
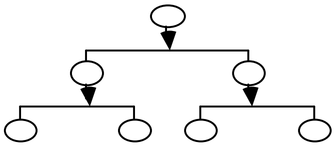
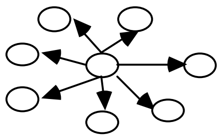
Metaphorical Graph	Brief Description
	Chain of components as in a *machine.*
	Network or circuit of components as in a *circuit.*
	Hierarchy of components as in an *organism.*
	Cluster of components as in a *society.*

Figure 3 - Some Simple Metaphorical Graphs and Associated Systemic Metaphors

In order to illustrate this we describe an analysis which looks at the systemic nature of the system (in this case an ecosystem), in the real world. Metaphorical analysis of this small piece of text can reveal a large amount of detail about the knowledge of the domain. As such, we are not primarily seeking to code the information into some kind of symbolic representation such as a semantic network. The focus of analysis is the characterisation of the domain, the choice of the post-analysis representational formalism follows from this.

> "An ecological system has a richly detailed budget of inputs and outputs...An ecosystem as we use the term, is a basic functional unit of nature comprising both organisms and their nonliving environment ... The living and nonliving components interact among themselves and with each other; they influence each others properties and both are essential for the maintenance and development of the system ... An ecosystem (can) be visualised as a grouping of components ... linked together by food webs, flows of nutrients and flows of energy." Bormann & Likens (1970), p92.

Firstly, there are a lot of words which reference the systemic metaphor (see Figure 2), such as: “system”, “interact”, “influence each other”, “linked together”. “System” is a very common word in everyday language and we characterise the systemic metaphors by a set of basic properties (called systemic M-properties (see Paton *et al.*, 1991b)): interacting parts, organisation, collective behaviour and whole system functionality. These M-properties provide a knowledge engineer with a means of anticipating certain aspects of the nature of the domain. For example, given that the text does not make ideas about organisation explicit, the knowledge engineer may wish to probe an expert about this issue in a subsequent elicitation session.

In order to provide the widest discrimination between systems, the systemic M-properties are associated with a set of systemic metaphors which inherit all the properties given above and include machine, organism, society, circuit, game, text and culture. There are further properties which typify particular systemic metaphors in the language used when talking about them. In terms of the language used in the above extract, three systemic metaphors are identified:

Machine	input, output, functional, component;
Organism	maintenance and development (note as a pair rather than single terms);
Circuit	flows.

These are but a small number of ideas which are associated with the systemic metaphors. Given this list a knowledge engineer may subsequently anticipate further ideas in the domain such as:

Machine:	efficiency, process, goal, purpose, power.
Organism:	growth, organised complexity, level, adaptability, openness.
Circuit:	transfer, conduit, cycle, transferred thing(s), drain.

These can be used to structure further elicitation from the expert through the anticipation of further properties.

The M-properties of the systemic metaphors listed above, though not exclusive to a particular type, are associated most clearly with that type (see Paton, 1991). Another important feature of this domain is the use of economic language (i.e., "budget"). This link to another domain is also very useful in helping to establish the common ontology and likely paramorphic models that can be used. A final comment relates to the authors' use of 'visualised'. This relates to the circuit metaphor, and is common to many ecosystem models. Alternatives, which are not explicit in the text, but could be anticipated include a hierarchy of niches, chain of compartments or complex of factors. These are directly related to the metaphorical graphs of Figure 3.

Knowledge analysis must place a crucial focus on the metaphorical nature of the domain for it provides a knowledge engineer with insight into its ontology, functionality, organisation and language (Paton & Nwana, 1990b). Once we appreciate the metaphors being used to describe a domain as a whole, it may be easier to flush out the analogical details in the knowledge acquisition process.

3.3 Scientific Theories can have Isomorphic Forms

Particular kinds of theory recur in different areas of scientific knowledge. Put another way, certain domains are related to others in the kind of theory that produced them. In this section we describe two specific examples and also discuss how each is currently being applied to issues associated with ‘emergent computation’ in computer science. For example, Darden (1987) and Darden and Cain (1989) discuss the occurrence of “selection” theories in the life sciences. Beginning with Darwin's theory of natural selection they show how it has been applied to immunology (clonal selection theory) and to the neurobiology of the brain (including memory and operant conditioning). The success of its application is with problems concerned with adaptation. The implication for knowledge analysis is that if generic (and hence abstract) theory kinds can be identified, then their features may provide valuable information for characterising domains. For example, we may anticipate that selection theories should have:

- a set of a given entity type;
- set members which vary according to a particular property (P);
- an environment in which the entity type is found;
- a factor in the environment to which members react differentially due to their possession/non-possession of the property (P);
- differential benefits (both short- and long-term) according to the possession/non-possession of the property (P).

If we can identify the occurrence of a selection theory we may thus anticipate important features of a domain. The example above illustrates the impact of one type of theory in three substantial areas of knowledge in the life sciences. Furthermore, we can see the application of selection to programming methodologies with the development of genetic (Darwinian) algorithms (Wilson, 1989), neural networks (Anderson & Rosenfeld, 1988) and the immune system (Farmer *et al.*, 1986). These approaches share many common abstract features and specific models of one can be applied to another (eg immune system as a parallel distributed processing network (Vertosick & Kelly, 1989)).

Another interesting example is that of foraging theory as applied in population biology. Rothbart (1991) presents an analysis of the work of Stephens & Krebs (1986) who describe a model of an animal as an ecomomic comnsumer in which the three basic concepts of consumer choice theory (utility, income and price). In this case, we see how a common theoretical type can be applied across domains. Furthermore, ideas from foraging models among insects have been applied to new ideas in computing (e.g., Deneubourg *et al*, 1986).

4 Some Results from Investigations

In this section we highlight how some of these foundational details of our view of knowledge analysis have been applied to two domains of expertise. Human knowledge is structured using theory and is metaphorical. We have proposed in Paton & Nwana (1990a) a methodology for knowledge analysis called SAAGS which exploits such emergent ideas. This four stage iterative cyclical modelling process receives as input a loose specification and outputs a more comprehensive specification that provides the characterisation of the domain. This can be fed into the subsequent design and implementation stages of KBS development. The principal stages are:

1. SPECIFICATION by the production of a characterisation of the domain. The seven top level features (see section one) must be accounted for in this stage. Additional details

must also include relevant models as well as descriptive details such as the epistemic boundaries yielded by the analysis.

2. ANTICIPATION of the nature of the domain. The top-level features allow the knowledge engineer to anticipate in breadth. The breadth is related to the model of the nature of a domain described above. As such it includes likely metaphors and theoretical frameworks that are expected to be used together with relations to other domains and the domain's history. Anticipations of these characteristics will help structure and guide the domain-based knowledge acquisition process which then takes place.
3. ACQUISITION of knowledge. This includes: knowledge elicitation from experts and reviewing textbooks, manuals and other knowledge sources. The outcomes from the anticipation stage are used to structure acquisition in a way which relates to the emerging nature of the domain. As such acquisition is driven by domain-related concerns. (Please note: in the terminology of the SAAGS approach, acquisition has this restricted meaning).
4. GENERATION of models including the synthesis of all outputs from the ACQUISITION stage into a collection of models. These models drive the analysis forward and provide the explicit means for confirming, refuting or elaborating on anticipations. This stage is needed to clarify ouputs should an expert approach a domain in a way that was not anticipated.

The goal of SAAGS is to produce a domain specification that is coherent, comprehensive, consistent and relevant and the cycle continues until the goal is reached. Thus, knowledge acquisition is only one of four stages and is surrounded by stages concerned with the processing (analysis and synthesis of frameworks of knowledge). We believe these other three stages are essential if knowledge acquisition is going to produce a meaningful description of the domain.

4.1 The Approach in Practice

In practice, SAAGS has been used as follows. The knowledge engineer (KE) is presented with an input specification (e.g. request to produce a particular KBS). A decision is made whether it is sufficient. The first step is to precisely establish a working vocabulary for the domain of discourse. If the KE knows little about the domain, an initial simple *acquisition* phase is done, possibly involving no more than reading though the contents, index and abstracts of relevant papers/books listing the major words used. The KE then *generates* a preliminary list of concepts, metaphors, and hypothesises some relationships within the domain and with other domains. The *specification* is refined, identifying the main top level concepts in the domain and the KE then *anticipates* likely metaphors, theories and other domains to which it is related. It is appreciated that these anticipations do not provide the KE with sufficient knowledge to characterise the domain but rather helps him/her structure the acquisition process and provide certain expectations on the knowledge to be gained subsequently.

The KE now organises the first session with the expert with goals of establishing some kind of rapport with him/her, casting a wide net over the domain in order to appreciate its boundaries (as the expert sees it) through eliciting the domain's history, metaphor and theories. At this stage the questions posed to the expert are thus quite *grand tour*. For example, the expert will usually be asked to give a story of what he/she does, together with discussion about what is in the domain, what the domain does and if it is related to any other domains. It is important to note here how anticipations have already guided the KE to ask the sort of questions which elicit aspects of domains proposed including metaphor, theory, relations to other domains, history and so forth. This session thus uses both structured and unstructured techniques. The audio-tape record is transcribed and analysed. Primary (paper) models are *generated* from the analysis including a vocabulary of discourse, list of global metaphors, concept map of related domains, list of sortal types and properties/attributes and preliminary relations between the latter. At this stage, the anticipated details are compared with the generated models. The differences between them furnish further questions for the next session with the expert. The *specification* is again refined; emerging outputs should include: organisation of the different sortal types and their properties, relations between the latter and global metaphors, identification of key tasks (relating to the domain's functionality), relations between objects, relations with other domains. The emerging theories and metaphors should structure the generated models as well as the representation of the specification/characterisation and hence the domain. Any KE will bring assumptions with her/him to the knowledge acquisition process. SAAGS helps to control them by making them explicit, and turns them to good effect by exploiting them during acquisition.

The cycle iterates in this manner. After each iteration the cycle returns to *specification* but a great deal has been learnt and to some degree structured. The sessions with the expert become more structured as wrong anticipations are cleared up and details are sought. As LaFrance (1987) notes, such details usually include: categorisation of the expert's concepts in a hierarchy (use the laddered-grid technique), ascertaining attributes (use repertory grid type questions), questions to determine interrelationships (eg eliciting a causal model for concepts), questions seeking advice (to elicit expert's recommendations and strategies for dealing with certain conditions) and questions to validate information obtained (use cross checking questions eg., to clear up possible some wrong anticipations). Metaphorical graph techniques and imagery (mentioned earlier) are also used to provide further insights in to the theories/metaphors of the domain; they are also used in eliciting the expert's reasoning process, which may be intuitively based. The process ends when the KE and expert mutually agree on the goal specification.

4.2 Knowledge Analysis in a Software Installation Domain

With SAAGS at a very formative stage, we investigated the domain of Hewlett Packard UNIX (HP-UX) software installation. We found that it was crucial to commence with an adequate and sufficient specification of the domain. This may require a kind of acquisition from the client about what the domain is (note: *not* the artifact) and may form the first iterative cycle. Following on from this, decisions made about the emerging nature of the domain should be exploited to suggest the kinds of analytical techniques, elicitation methods and representational schemes to use.

The requirement for disciplined, theoretically relevant and controlled anticipations prevents time-wasting and establishes important expectations about the vocabulary of discourse at an early stage in a project. One example concerned the expert's use of the notion of the "system". Analysis of the interview text revealed that he used the term in three different ways. The nature of the discourse indicated an iconic aspect and so, at the next interview session, we decided to show him a set of simple metaphorical graphs (similar to those in the left-hand column of Figure 2). His choice allowed us to clarify the nature of the different systems he had described.

One important issue we faced could be called "the problem of representation", that is, how and why we represent outputs from the SAAGS stages in the manner we do? It was clear we should not directly use a representational medium such as frames, rules, semantic networks or predicate logic as it would drive or influence the analysis. This is a very important lesson to learn. A key criticism of current knowledge acquisition tools is that they are governed by representational concerns. We argue that it should be the nature of the domain that should guide both the knowledge acquisition process and the final form of the knowledge base (see also Woodward, 1989).

Other lessons were learned. For example, we found that using SAAGS allowed us to appreciate how easy it is to become entangled with the detailed concerns of knowledge representation formalisms rather than focussing on the domain knowledge as a whole (related to that mentioned earlier). We saw that the text required more than a lexical analysis. All these lessons contributed to the enhancement of the approach.

4.3 The Analysis of Knowledge in a Neural Networks Domain

A subsequent project which applied the SAAGS approach concerned the application of neural networks to industrial problems. The goal of the analysis was to identify what type of neural network architecture is most suitable for a given type of application. The approach as described in Section 4.1 was followed and it is important to report how it disciplined, more than in the software installation exercise, our actions and analyses. It would be lengthy and inappropriate to include the detailed results of the work (for further information see, Nwana & Paton, 1990, Nwana *et al.*, 1990 & Paton *et al.*, 1991c). A discussion relating to the foundational ideas of this paper is provided instead.

The exercise more clearly vindicated the importance of some issues including metaphor, theory, relations to other domains and history. During the *grand tour* session, the expert provided us with his appreciation of the domain's history, primarily in physics, statistics and Operational Research. At the generation stage we were able to compare the expert's appreciation with our more general model, based on textbook and journal analyses. This gave us valuable information on limitations of neural network applications to this specific problem domain. Other domains which he noted as relevant to his knowledge included neurophysiology, psychology and mathematics. This helped to clarify details of network structure and functionality. It is important to identify such details early in a SAAGS investigation since it is essential to detect what theories underpin an expert's understanding, so that fruitful anticipations can be made. By the end of the first elicitation session we had obtained a broad view of the domain in terms of basic structural details, purpose, relations to other domains. We also found how important the use of metaphorical thinking was, especially as the domain has no common theory (in the sense of a shared body of knowledge). In this case we were able to identify the use made of some metaphors; for example, the expert's use of the machine metaphor.

An appreciation of systemic M-properties (see Figure 2) helped us to anticipate and characterise systemic and machine concepts. As such we were able to begin the second iterative cycle with the expert with certain hypotheses which needed further investigation. SAAGS provided us with a disciplined framework for identifying this way of thinking. This enabled us to probe the expert for specific details, such as the machines parts and interrelations, organisation and functionality (systemic M-properties) as well as inputs, outputs, mechanism and optimisation (machine M-properties). We also found that when talking about the behaviour of a network (in addition to its structure) the preferred metaphor was that of an organism and in this case ideas such as learning, adaptability and babies-as-untrained networks were used. When also probed to explain the details of how neural networks learn, he also resorted to using landscape metaphors, in this case boulders rolling down a hill and getting into ‘local minima’.

A lot of detail had been analysed by the end of the second elicitation session with the expert and it was becoming clear that the domain had explanatory power. Evidence for this was deduced from the language of the expert's response. Specifically, ideas associated with network configuration, dimensionality and “architecture” were abstractions described in terms of other network components (such as the processing units and topology). However, no iconic details were made explicit. The SAAGS approach requires that this kind of information is made available because it allows models, such as those associated with inference mechanisms, to be described. The need for this kind of informationwas therefore anticipated. Part of the next iterative cycle sought to elicit information about underlying theory through a set of focussed interactive questions. What emerged was very valuable. The expert noted:

> "There are two different attitudes to the neurophysiology analogy...one is that birds, insects and bats fly. But when the Wright brothers designed their first aeroplane the wings did not flap and it flew. The moral of the story is you don't have to do things the way the biological systems do it in order to be efficient. And the other attitude is that non-biological systems work very poorly. If you can't do something by using a physical system, you can mimic it, i.e., the biological system. I tend to lean to the second view and that's entirely a personal preference, its a question of background. My background is in physiology rather than physics."

The above text provides a knowledge engineer with the basis of an explanatory model for this domain. Could this kind of information have been acquired by other means? The value of the SAAGS approach is that it requires that this kind of information <u>is</u> acquired. How does it assist in the model of the domain? To some workers in the neural networks community the response may seem obvious. For people building rule based KBSs it will not be necessary. For SAAGS analysis it is crucial: it underpins the explanatory framework. The value of such information is that it provides a deeper understanding of the domain and also points to the analogies the expert may use when solving problems in the domain such as seeking to apply particular architectures to novel situations. The discussion above provides the context and depth of characterisation for a rule which the expert provided for deciding on the application of an artificial neural network to a problem:

IF	It is difficult to elicit knowledge from the expert because expert cannot tell you the rules, for example it is too basic (eg. smell; what are the relevant rules to smell or see?)
OR IF	
	rules are extremely many to be tractable (ie complex)
OR IF	
	the problem is non-linear
THEN	
	Use Artificial Neural Networks Approach.

5 Concluding Comments

We have attempted to highlight some of the foundational ideas needed to provide sufficient depth to analyse a domain. The central issue in knowledge analysis is the characterisation of a domain. It is only after such an exercise that design and implementation of a knowledge-based system should proceed. The foundational details we have described give a philosophical and conceptual basis for knowledge acquisition. We describe how the ideas have been exploited in a methodology called SAAGS which, in turn, has been applied experimentally to two domains with so far interesting and promising results. Further consolidation of this work is in progress with the analysis of other non-trivial industrial domains.

Acknowledgements
Thanks to our experts Ken Chan and Brian Ward and to Professor Rom Harré for his advice and comments, especially in relation to section 3. The MEKAS Project is sponsored by Shell Research Ltd (Thornton Research Centre) and Unilever Research (Port Sunlight Laboratory).

References

Alexander, J. H., Freiling, M. J., Shulman, S. J., Staley, J. L., Rehfus, S. & Messick, S. L. (1986), "Knowledge level Engineering: Ontological Analysis", *Proceedings of AAAI* **5**, Philadelphia, 963-968.

Aronson, J. L. (1984), *A Realist Philosophy of Science*, London: Macmillan

Aronson, J.L. (1991, in press), "Verisimilitude and Type Hierarchies", To appear in *Philosophical Topics*.

Bhaskar, R. (1978), *A Realist Theory of Science,* Sussex: Harvester Press.

Black, M. (1979), "More about Metaphor" In Ortony, A. (ed.), *Metaphor and Thought*, London: Cambridge University Press, 19-43.

Bormann, F.H. & Likens, G.E. (1970), "Nutrient Cycles of an Ecosystem", *Scientific American*, **223**(4), 92 - 101.

Boose, J.H. (1989), "A Survey of Knowledge Acquisition Techniques and Tools", *Knowledge Acquisition*, **1**, 3 - 37.

Boyd, R. (1979), "Metaphor and Theory Change: What is 'Metaphor' a Metaphor for?", In Ortony, A. (ed.), *Metaphor and Thought*, London: Cambridge University Press, 356-408.

Bradshaw, J. & Woodward, B. (1989), "Knowledge Acquisition Tools", discussion in Boose, J.& Gaines, B. (eds.), *Knowledge Acquistion for KBS Newsletter* , **1**(3), 12.

Buchanan, B. G., Barstow, D., Bechtal, R., Bennett, J., Clancey, W. J., Kulukowski, C., Mitchell, T. & Waterman, D. A. (1983), "Constructing an Expert System", In Hayes-Roth, F., Waterman, D. A. & Lenat, D. B (eds.), *Building Expert Systems*, Reading, Mass.: Addison-Wesley.

Carey, S. (1985), *Conceptual Change in Childhood*, Mass.: MIT Press.

Chandrasekaran, B. (1985), "Generic tasks in Knowledge Based Reasoning: Characterizing and designing Expert Systems at the 'Right' level of Abstraction", *Proceedings of the IEEE 2nd Conference on Artificial Intelligence Applications*, 294-300.

Clancey, W. J. (1986), "Heuristic Classification", In Kowalik, J. S. (ed.), *Knowledge based Problem Solving*, Englewood Cliffs, N. J.: Prentice Hall, 1-67.

Clancey, W. J (1989), "The Knowledge Level Reinterpreted: Modeling How Systems Interact", *Machine Learning*, **4**, 285-291.

Darden, L. (1987), "Viewing the History of Science as Compiled Hindsight", *AI Magazine*, Summer, 33-41.

Darden, L. & Cain, J. A. (1989), "Selection Type Theories", *Philosophy of Science*, **56**(1), 106-129.

Deneubourg, J. L., Aron, S., Goss, S., Pasteels, J.M.& Duerinck, G. (1986), "Random Behaviour, Amplification Processes and the Number of Participants: How they Contribute to the Foraging Properties of Ants", *Physica*, **22D**, 176-186.

Deutsch, K. (1951), "Mechanism, Organism and Society: Some Models in the Natural and Social Sciences", *Philosophy of Science*, **18**, 230-252.

Downs, E., Clare, P. & Coe, I. (1988), *Structured Systems Analysis and Design Method: Application and Context*, Hemel Hempstead, Herts: Prentice Hall.

Engler, G. (1990), "Aesthetics in Science and Art", *British Journal of Aesthetics* **30**(1), 24-33.

Farmer, J. D., Packard, N. H. & Perelson, A. S. (1986), "The Immune System, Adaptation and Machine Learning", *Physica*, **22D**, 187-204.
Finch, I. (1989), "Explanation systems in engineering domains", *Proceedings of the International conference on systems in engineering*, Wuhan, China.
Forrester, J. W. (1968), *Principles of Systems*, Mass.: Wright Allen.
Gaines, B.R. (1989), "Social and cognitive processes in knowledge acquisition", Knowledge Acquisition, **1**, 39 - 58.
Genter, D. & Gruding, G. (1985), "The Evolution of Mental Metaphors in Psychology: A ninety year perspective", *American Psychologist*, **40** (2), 181-192.
Guida, G. & Tasso, C. (1989), "Building expert systems: From Life Cycle to development methodology", In Guida, G. & Tasso, C. (eds.), *Topics in Expert Systems Design, Methodologies and Tools*, Amsterdam: North Holland, 3-24.
Harré, R. (1970), *The Principles of Scientific Thinking*, London: Macmillan.
Harré, R. (1986), *Varieties of Realism*, Oxford: Blackwell.
Hesse, M.B. (1963), *Models and Analogies in Science*, London: Sheed & Ward.
Hughes, S. (1986), "Question classification in rule-based systems", In Bramer, M. A. (ed.), *Research and Development in Expert Systems III*, Cambridge: University Press, 123-131.
Kangassalo, H. (1990), "Foundations of Conceptual Modelling: A Theory Construction View", In Kangassalo, H., Ohsuga, S. & Jaakkola, H. (eds.), *Information Modelling and Knowledge Bases*, Amsterdam: IOS, 19-35.
Karmiloff-Smith, A. (1988), "The Child is a Theoretician, not an Inductivist", *Mind and Language*, **3**, 183-195.
Keil, F. C. (1989), *Concepts, Kinds and Cognitive Development*, London: MIT Press.
Kidd, A. (1987), "Knowledge Acquisition: An Introductory Framework", In Kidd, A. (ed.), *Knowledge Acquisition for Expert Systems: A Practical Handbook*, London: Plenum, 1-15.
Kitto, C. M. & Boose, J. H. (1987), "Heuristics for expertise transfer: an implementation of a dialogue manager for knowledge acquisition", *International Journal of Man-Machine Studies*, **26**, 183-202.
Kuhn, T. S. (1979), "Metaphor in Science", In Ortony, A. (ed.), *Metaphor and Thought*, London: Cambridge University Press, 409-419.
Lackoff, G. & Johnson, M. (1980), *Metaphors we live by*, Chicago: University Press.
LaFrance, M. (1987), "The Knowledge Acquisition Grid: a method for training knowledge engineers", *International Journal of Man-Machine Studies* **26**, 245-255.
Lehman, R.S. (1977), *Computer Modelling and Simulation*, New York: Lawrence Erlbaum.
Lewis, T.G. & Smith, B.J. (1979), *Computer Principles of Modelling and Simulation*, Boston: Houghton Mifflin Co.
Medin, D. L. & Wattenmaker, W. D. (1987), "Category cohesiveness, theories and cognitive archeology", in Neisser, U. (ed.), *Concepts and Conceptual Development*, Cambridge: University Press, 25-62.
Morgan, G (1980), "Paradigms, Metaphors and Puzzle Solving in Organisation Theory", *Administrative Science Quarterly*, **25**, 605-622.
Newell, A. (1982), "The knowledge level", *Artificial Intelligence*, **18**, 87-127.
Nwana, H.S. & Paton, R.C. (1990), "Report of a Case Study on the Application of the SAAGS Approach to a Domain Concerned with Neural Networks", MEKAS Report Number 12, The Department of Computer Science, Liverpool University, 15th November, 1990.
Nwana, H.S., Paton, R.C., Shave, M. J. R. & Bench-Capon, T. J. M. (1990), "The Relevance of Domain Characterisation to Knowledge Acquisition for Second Generation Expert Systems", MEKAS Report Number 17, Department of Computer Science, The University of Liverpool, December 19. Submitted for publication.

O'Neill, M. & Morris, A. (1989), “Expert systems in the United Kingdom: an evaluation of development methodologies”, *Expert Systems*, **6**(2), 90-98.

Paton, R.C. (1991), “Towards a Metaphorical Biology”, Working Paper Series: CS/1991/21, Department of Computer Science, The University of Liverpool. Submitted for publication.

Paton, R. C. & Nwana, H. S. (1990a), “Domain Characterisation through Knowledge Analysis”, *Proceedings of AAAI-90 Workshop on Knowledge Acquisition: Practical Tools and Techniques*, Boston, Massachussetts, July.

Paton, R. C. & Nwana, H. S. (1990b), “Metaphorical Aspects of Thinking in the Social Sciences”, MEKAS Report Number 6, Department of Computer Science, The University of Liverpool.

Paton, R. C., Nwana, H. S., Shave, M. J. R., Bench-Capon, T. J. M. & Hughes, S. (1991a) “Transfer of Natural Metaphors to Parallel Problem Solving Applications”, in Schwefel, H-P & Maenner, R. (eds.), *Parallel Problem Solving from Nature*, *Lecture Notes in Computer Science*, Berlin: Springer-Verlag, 355-365.

Paton, R. C., Nwana, H. S., Shave, M. J. R. & Bench-Capon, T. J. M. (1991b), “Systemic Metaphor Analysis for Knowledge Based System Development”, to appear in Jackson, M.C. et al (eds.), *Systems Thinking in Europe*, London: Plenum.

Pepper, S. (1928), “Philosophy and Metaphor”, *Journal of Philosophy*, 130-132.

Recogzei,S. & Plantinga, E.P.O. (1987), “Creating the domain of discourse: ontology and inventory”, *International Journal of Man-Machine Studies*, **27**, 235 - 251.

Rothbart, D.: 1984, “The Semantics of Metaphor and the Structure of Science”, *Philosophy of Science*, **51**, 595-615.

Rothbart, D. (1991), “Metaphors, Models and Meanings in Science”, paper to be presented at a seminar on metaphor, Professor Rom Harré, Director, Oxford University, June 6th.

Shapere, D. (1977), “Scientific Theories and their Domains”, In Suppe, F. (ed.), *The Structure of Scientific Theories*, Urbana: Illinois University Press, 518-599.

Shaw, M. L. G. (1989), “A Grid-Based Tool for Knowledge Acquisition”, *Proceedings of IJCAI-1989 Workshop on Knowledge Acquisition*, Detroit, Michigan, 19-22.

Soskice, J.M.: 1986, *Metaphor and Religious Language*, Oxford: Clarendon.

Stephens, D.W. & Krebs, J.R. (1986), *Foraging Theory*, New Jersey : Princeton.

Vertosick, F. T. & Kelly, R. H. (1989), “Immune network theory: a role for parallel distributed processing”, *Immunology*, **66**, 1-7.

Watson, I. D., Shave, M. J. & Moralee, D. S. (1989), “A knowledge analysis methodology using an intermediate knowledge representation based on conceptual graphs”, Proceedings of the 9th Int. Workshop on Expert Systems and their Applications **1**, Avignon, France, 183-198.

Way, E.C. (1991), *Metaphor and Knowledge Representation*, Dordrecht: Kluwer.

Wielinga, B. J., Bredeweg, B. & Breuker, J. A. (1988), “Knowledge Acquisition for Expert Systems”, *Proceedings of ACAI-88*.

Wilson, S. W. (1989), “The Genetic Algorithm and Simulated Evolution”, In Langton, C. G. (ed) (1989), *Artificial Life* **VI,** California: Addison-Wesley, 159-165.

Woodward, B. (1989), “General Structures as Domain Models to Guide Knowledge Acquisition”, *Proceedings of IJCAI-1989 Workshop on Knowledge Acquisition*, Detroit, Michigan, 10-14.