IEEE Transactions on Systems, Man, and Cybernetics Epistemological Models, Medical Knowledge Based Systems

An Epistemological Framework for Medical Knowledge Based Systems

Marco Ramoni

Mario Stefanelli
Dipartimento di Informatica e Sistemistica
Università di Pavia, Italy Lorenzo Magnani
Dipartimento di Filosofia
Università di Pavia, Italy Giovanni Barosi
Dipartimento di Medicina Interna
IRCCS S. Matteo, Pavia, Italy

January 8, 2000

Abstract

There is a general agreement about the need of identifying a high-level conceptual design in knowledge based systems that is behind the implementation. A number of perspectives for understanding knowledge based systems at this high level has been proposed. Major examples of these perspectives are based on the concept of Heuristic Classification, Deep-Shallow Systems, Problem Solving Method, and Generic Task. Each of these ideas focuses on a particular feature of reasoning: the inference structure, the models of domains knowledge, the sequence of actions need to solve a problem, and the task features.

This paper presents a new abstraction paradigm aiming at unifying these different perspectives. The proposed model should be able to account for all of the conceptual features of knowledge based systems, thus making clear which features are intrinsic to the problem and which are artifacts of the implementation. Our proposal is therefore based on a two levels analysis of knowledge based systems: an epistemological and a computational level. At the first level, ontology and inference model of a knowledge based system have to be defined. Ontology represents the conceptual model of domain knowledge, while the inference model is the conceptual representation of the inference structure needed to solve a problem or to execute a task by managing that ontology. At the computational level, methods and formalisms should be adopted after the epistemological analysis has been carried

out and taking into account the constraints which derive from the conceptual structure of the domain knowledge, patterns of inference, and tasks to be executed.

We constrained our study to medicine and identified three generic tasks: diagnosis, therapy planning, and monitoring. The main result of this analysis is that these generic tasks manage different ontologies, but can be executed exploiting a unique inference model. Such a model involves three different inference types (abduction, deduction, and induction) and is described in some details. Finally, computational issues are discussed to argue that the present model provides a conceptual view on existing systems, and some design insights for future ones.

1 Introduction

Much of the earlier work in Artificial Intelligence (AI) occurred in the very active medical AI community of the 1970s. The best-known medical knowledge-based systems (KBS) of this era were PIP [1], CASNET [2], MYCIN [3], and INTERNIST-1 [4]. Although these KBS used different methods and formalisms for representing and managing knowledge, none of them exploited any genuine pathophysiological theory but only *empirical* associations between manifestations and diseases [5]: the elements of medical knowledge embedded in their knowledge bases were nothing more than *pragmatic constructs* [6] codifying previously practical experiences of domain experts.

The opinion that these systems had serious limitations has begun to arise in the early of 1980s [7]. Namely, they were often judged as *brittle*, since they exhibited a sudden performance degradation when the problem at hand was near or beyond the limits of their domain knowledge (i.e. when they did not contain any piece of knowledge covering exactly the question they were asked about) [8]. A graceful performance degradation — as it happens with human experts [9,10] — would be more desirable. Moreover, these systems were not able to generate satisfactory explanations [11,12], since they could only show the pragmatic constructs composing their own knowledge bases. In addition, they were unable to solve "novel" problems, i.e. problems which the designer of the KBS did not anticipate while building the system [5]. Finally, there were several difficulties in acquiring and maintaining their knowledge bases, because the systems needed the heuristic insight of a domain expert more than mere domain knowledge. Hence, knowledge acquisition has come to be seen as the bottleneck in the knowledge engineering process [13], since

3

these pragmatic constructs involved tacit assumptions and nonarticulable skills whose acquisition was quite difficult.

Several researchers [13,14,15] claim that solutions for these problems may be achieved through a suitable design of the conceptual architecture before starting any implementation activity. This level of design has been called *knowledge level* [16], and it has been clearly separated from the *computational level*. The knowledge level focuses on the structures and contents of knowledge and it is independent to the implementation, while the computational level deals with methods and formalisms exploited to implement a working system.

At the knowledge level, four main approaches have been proposed during the last decade in order to provide a theoretical foundation of KBS [15]. These approaches are based, respectively, on the concepts of *Heuristic Classification* [17], *Deep-Shallow system* [18], *Problem-Solving Method* [19,20], and *Generic Task* [21]. They are usually regarded as *abstraction paradigms* leading to *rational reconstructions* of KBS. However, their formulation does not represent only a theoretical problem but also a fundamental step in KBS design.

Heuristic Classification provides a theoretical framework for KBS stressing features and properties of heuristic reasoning (*experiential reasoning* in the Heuristic Classification's lexicon) that a KBS should develop. As an abstraction paradigm, heuristic classification focuses on the *inference structure* underlying expertise. There are several advantages in an analysis focusing on inference structures: it allows to identify basic components of heuristic pathways by highlighting features that are beyond the domain knowledge and, therefore, it shows similarities and differences across different problem types and application domains. Systems such as NEOMYCIN [22] and GUIDON [23] have been developed following the heuristic classification abstraction paradigm.

Another abstraction paradigm — based on the dicotomy Deep-Shallow Systems — focuses on the theoretical structure and contents of domain knowledge, instead of inference structure. A system is said to be *deep* when the problem solver works on an explicit symbolic representation of the structure and behavior of the underlying pathophysiological system, while this type of knowledge is only implicitly represented in a shallow system. Following earlier remarks made by Davis [5], several authors pointed out that the domain knowledge becomes explicit and more accessible when the inference structure is separated as much as possible from the domain knowledge. However, in order to exploit domain knowledge in a suitable way, a

4

domain-independent inference structure needs to be designed. As Levesque [24] pointed out, representing explicitly and separately domain knowledge and inference structure leads to assume that there is a calculus of thought manipulating knowledge independent from its contents. Thus the definition of the inference structure turns out as a fundamental task also for the Deep-Shallow system approach.

The paradigm based on the concept of Problem Solving Method is a knowledge level characterization of the sequence of actions that enables the agent to execute a certain task in a specific domain. A problem solving method is usually defined by some mechanisms (1) to generate a set of candidate actions and (2) to select among these candidates the action to be executed. For instance, diagnosis [25] can be viewed as a process of covering and differentiating (i.e. (1) find possible diseases covering most symptoms and (2) differentiate between he remaining explanations) and construction [20] as a process of proposing and revising solutions (i.e. (1) propose a partial solution and (2) revise solution by resolving violated constraints). One of the most interesting results of this approach is the claim that each of these problem solving methods needs to fill domain knowledge into some method-specific roles: diagnosis, for instance, requires knowledge about the relationships linking symptoms to possible diagnosis and additional symptoms to further differentiate. Hence, domain knowledge may be no longer regarded as independent from its concrete use in the problem solving process, as in deep systems. However, domain knowledge is still explicitly and separately represented and this allows us to gain the advantages of maintenability and systematicness of domain knowledge.

The basic idea underlying Generic Tasks approach is that every real world complex task may be decomposed into simpler subtasks, having input/output relations between them. Each task falls into major classes of tasks, named *generic tasks*. Chandrasekaran identified a small set of generic tasks, such as interpretation, classification, diagnosis, design, and so forth. These generic tasks represent basic elements in the conceptual architecture of a KBS, and they show similarities across application domains. MDX [26] was the first medical KBS developed within the generic task framework.

Besides their differences, all these abstraction paradigms, Heuristic Classification, Deep-Shallow Systems, Problem Solving Methods, and Generic Tasks, are not alternative each other. Clancey [17] showed how heuristic classification may be useful in building a deep system as GUIDON; Chandrasekaran [27] stressed that the generic task approach may allow us to develop a deep system and that the concept of generic task may involve

5

the definition of a suitable inference structure [28]. MDX-II provides an example of a deep system developed following the generic task approach. On the other hand, a *deep system* such as NEOCRIB [29] has been described in terms of the generic tasks it executes. Finally, domain knowledge is explicitly and separately represented in MOLE [25], a tool for developing KBS based on the cover-and-differentiate method, that is a specific problem solving method.

This paper presents a new abstraction paradigm which starts from the well-known *knowledge/computational level* distinction and tries to exploit inference structure, domain knowledge and task features in order to provide a unified knowledge level analysis of KBS. The basic idea is to provide an epistemological framework able to subsume and integrate these knowledge level approaches. Therefore, we replaced it with the term *epistemological level*. Conceptual features of tasks, inference structure and domain models fall into this level and the dichotomy *knowledge/computational level* becomes the dichotomy *epistemological/computational level*.

The reminder of this paper is structured as follows. Section 2 will be devoted to develop our unified epistemological analysis of KBS components. Then, we will argue that a unique inference structure, called *Select and Test Model*, underlies medical reasoning. Finally, we will discuss some computational issues in order to show that the present model provides a conceptual view on existing systems, and some design insights for building future ones.

2   The Epistemological Level

A KBS contains two main types of knowledge: knowledge about the domain and knowledge about inference procedures which have to be exploited to solve a problem. An epistemological analysis focuses on the conceptual features of these high-level components of KBS, that is on the definition of the underlying *ontology* and *inference model*. The result of including these two components at the epistemological level is outlined in Figure 1. Ontology represents the conceptual model of entities and relationships composing the domain knowledge, while the inference model is the conceptual representation of the inference structure employed to execute a task by managing that ontology. Abduction, deduction and induction represent the basic elements of the inference model of medical reasoning described in the next section.

The computational level has been broken down into two parts: methods and formalisms. This outlines one of the major issue addressed by this

6

fig1.eps3.5in

Figure 1: A schematic representation of the two level analysis of Knowledge Based Systems and elements belonging to each level.

paper: different disciplines (such as logic, mathematics, decision theory, and AI) may provide suitable formalisms for building KBS able to solve real medical problems.

In medical reasoning, however, the task to be executed is actually regarded as an aggregate of sub-tasks pursuing different goals. Several researchers [30,31] argued that the main goal of medical reasoning is to modify the current abnormal state of affairs in a patient, that is to suggest therapeutic treatments or, at least, prognostic forecast given a specific disease and/or treatment, rather than provide insight. This very tough goal requires the execution of the following three tasks: diagnosis, therapy planning and monitoring [32,33,34]. We claim that these basic tasks manage different ontologies, but they can be executed exploiting a unique inference model.

2.1 Ontology

Contemporary analytical philosophy calls ontology the conceptual definition and cataloguing of entities and relationships existing in the world, or, at least, in our ways of conceiving the world. It has been widely investigated by Strawson [35] who developed a general theory of conceptual schemas of conceiving worlds. This theory has been shown to be useful in understanding much of the work in medical KBS [36]. Several researchers pointed out the important role of ontology in different AI domains, such as qualitative reasoning [37], non monotonic reasoning [38], knowledge-base theory [39,40].

2.1.1 Entities and Relationships

An ontology allows us to conceive a unified abstract scheme where things, states, actions, causes and events may be defined, enumerated, and represented as generic terms. This property fulfills the requirements of an epistemological analysis of a KBS and ontology may be regarded as the conceptual model of the embedded domain knowledge.

We distinguish two basic classes of elements in an ontology: *entities* and *relationships* between entities. For instance, findings, diseases, and therapies

7

may be entities in the ontology of a medical KBS, while causality and type-subtype may be relationships.

The separation between ontology and inference model should allow us to represent ontology in a KBS (i.e. *KBS ontology*) as well as it is organized in the scientific medical knowledge (i.e. *source ontology*). Following McCarthy and Hayes [41], we call this feature "representational adequacy". This representational adequacy allows us to represent genuine theory of diseases, when available, or, at least, to deeply exploit features of the available domain knowledge in order to generate satisfactory explanations, to solve unexpected problems, and to simplify the process of acquiring and maintaining domain knowledge.

In this sense we claim that first generation KBS exploited no real theories of diseases. Indeed, pragmatic constructs mapping manifestations into disease entities loose their original organization in the scientific medical knowledge (as it is set out in medical textbooks). The need for an efficient exploitation of medical knowledge often forced KBS developers to represent domain knowledge according to the conceptual limits imposed by the inference processes to be applied to execute a task. In such a case, ontology is said to be compiled [18,42] into an inference model, that is it becomes implicit. The results of such a compilation are pragmatic constructs.

2.1.2 KBS Ontologies

Several studies [4,6] have pointed out that there are two main ways of organizing medical knowledge, that is, there are two main kinds of source ontology: a causal and a taxonomic ontology. We are interested in analysing how these source ontologies can be translated into a KBS ontology.

Causal ontologies. According to several authors, a causal ontology is a definitional character of deep systems [7,43]. Unfortunately, exactly what constitutes causal knowledge and, therefore, a causal ontology was never made clear. Philosophical tradition has not synthesized any coherent and widely accepted definition of causality and, as Karp and Wilkins [12] pointed out, when AI authors provide few clues to what they believe causality is, it is difficult to accept any claim about its lack in some KBS. Nevertheless, several systems have been developed to represent some kind of causal ontology and we will try to investigate features of a causal ontology analysing how it has been represented in KBS.

8

There are at least two alternative ways of representing a causal ontology [6,44]. Firstly, they have been represented as a network, where nodes represent states and events occurring in the patient and links a causal relation between them. This kind of causal ontology has been exploited in some earlier medical KBS, such as CASNET and CADUCEUS [45]. Simon [6] stressed the "qualitative" nature of these representations, that seems to be closest to the clinical way of thinking than usual quantitative models provided by mathematics. But such an approach appears too weak to represent a genuine biomedical theory because it relates state to state and event to event, without describing structure and laws of behavior of the the underlying pathophysiological systems. This causal ontology is not able to provide both a description of the the normal behavior and the evolution from normal to abnormal behavior under the action of a given pathogenetic disturbance. This approach to causal ontology has been strongly enhanced by Patil developing ABEL [46]: diseases and pathophysiological states are represented at different levels of detail. Some of the highest level nodes may be expressed at next level in greater detail. This effort has been motivated by the computational advantages given by the possibility of exploiting multi-level description of a pathophysiological system, if needed, but the real theoretical challenge here was to represent a low-level description of the system and, therefore, genuine pathophysiological theories. Unfortunately, also ABEL's pathophysiological level is constituted by a network of states and not by a description of structure and behavior of the system. The main difference between knowledge represented at the clinical level and that represented at lower levels lies in the *granularity* of knowledge rather than in their ontological features.

An alternative way of representing a causal ontology should describe structure and behavior of underlying systems. This approach has been widely exploited for diagnosing faults in circuits and physical devices and it seems to be closer to the current AI interpretation of causality than the state-to-state approach. In this approach, the causal relation links variables rather than states. These variables represent attribute of the system to be modeled, and a causal relation represents the fact that a change of a variable produces a change in another variable. By describing the structure and the behavior of a system, this approach would allow us to represent genuine would genuine biomedical theories as proposed in medical literature: pathophysiological situation can be interpreted as the effect of some perturbations occurring in some system variables. However, in a real clinical setting, physicians usually lack of a complete description of the internal structure of the

9

patient. Nonetheless, physicians do appear to be able to reason with models of structure and behavior of the system [47] since they use *qualitative models* [48,49,47]. This qualitative approach has been applied to model some pathophysiological systems in medical KBS using QSIM [44,50]. QSIM [47] provides a descriptive language, consisting of qualitative constraints that abstract the relationships in a differential equation, to represent the structure of a pathophysiological system and a simulation algorithm to infer its qualitative behavior. Moreover, from the ontological point of view, qualitative modelling seems to be more adequate to represent genuine biomedical theories, since it captures both the sistematicness of a model of structure and behavior and the qualitative attitude of reasoning showed by physician.

Taxonomic ontologies. Unfortunately, in many medical domains these models of structure and these laws of behavior are still unknown or, at least, incomplete and controversial. In such a case, the source ontology is organized in a taxonomic way; disease entities are hierarchically set out and empirically defined as clusters of manifestations [4,6], and no knowledge of underlying mechanism is available. These ontologies are typically represented as prototypes rather than as conditions for transition from a state to another [51]. This kind of ontology is defined as "taxonomic" because the relation of subsumption is the leading one: a concept (for example, a disease entity) is defined as a cluster of attributes (for example, patient's manifestations) and it is placed above those concepts that it subsumes (i.e. the set of manifestations defining the current disease entity is superset of those sets defining subsumed disease entities) and below those that subsume it (i.e. the set of manifestations defining the current disease entity is subset of those sets defining disease entities which subsume it). This definition may occur with exceptions, so justifying much of the work about inheritance theories. An example of taxonomic KBS ontology in medical domain is the prototypical approach followed in developing CENTAUR [52].

Notice that these definitions of concepts embodied in taxonomic ontologies are *empirical* but not *pragmatic* constructs: even though they are based on the surface mapping of clusters of manifestations into disease entities there is no inference purpose in their organization (i.e. the inference model is not compiled into them). According to the previous definition of deepshallow systems, the presence of a taxonomic ontology in a KBS does not lead necessarily to define it as a *shallow system*. Since deep systems have been defined as the result of separating the ontology from the inference

model and of explicitly representing both of them, a deep system may be built even by exploiting a taxonomic ontology. This avoids to exclude deep systems from a wide part of medicine, where knowledge about structure and behavior has not yet been made available. On the other hand, those KBS which tried to exploit some kind of causal ontology, such as CASNET, can be considered as shallow systems, because their ontology and inference model were not fully separated. Therefore, in our opinion, *a causal ontology is neither a necessary nor a sufficient condition to build up a deep system.*

Neither of these two ontological views may be a priori preferred because KBS have to deal with a wide range of medical domains. Moreover, as Shaffner [53] pointed out, these two types of organization usually overlap in the source ontology of most medical domains and the expressive power of KBS ontologies should be measured by their ability to mirror this overlap as much as possible. CADUCEUS, for instance, integrates causal and taxonomic ontologies (called "nosological models"). The same approach has been followed by NEOANEMIA [54].

But an ontology is dumb. It is a conceptual scheme underlying domain knowledge and providing only a static structure of entities and relationships. Therefore, it is not able by itself to execute any task, to draw inferences about the world or to make decisions.

2.2 Inference Model

As we have outlined above, there are two main kinds of knowledge embedded in a KBS: one concerns the world and the other concerns procedures able to manage the available knowledge about the world in order to solve problems. As ontology is the conceptual model of the first kind of knowledge, an *inference model* can be viewed as a conceptual model of the second one.

2.2.1 Inference

An inference is generally defined as the process of deriving true statements (called *conclusions*) from other true statements (called *premises*). It is usually a multistep process (i.e. we often need to derive intermediate conclusions). Each of these steps has to be sanctioned by an *inference type*, representing a general class of procedures taking these premises as inputs and leading to conclusions. Inference types are building blocks for an inference model.

11

2.2.2 Inference Types

According to Peirce [55], three main *inference types* can be distinguished: *deduction*, *abduction*, and *induction*. They have been defined by Peirce as fundamental and prelogic characters of reasoning. For our purpose, these three inference types may be defined as follows.

Deduction. Deduction is the inference type allowing one to derive a statement, given some *general laws* (or general rules, in Peirce's lexicon), and some singular statements called *initial conditions*. Knowing the general law "All patient affected by acute hepatitis manifest an increased value of serum transaminases" and the singular statement "John is affected by acute hepatitis", one can *deduce* that "John manifests an increased value of serum transaminases".

Notice that this definition of deduction is independent from the ontology we choose. The general law may be viewed as a prototypical property of acute hepatitis or as a causal relation between the state "The patient is affected by acute hepatitis" and the state "The patient manifests an increased value of serum transaminases". The general law can also be viewed as "laws of behavior", and the singular statement as the initial situation of the system under analysis. This ontology-independence is a common features of all inference types.

Abduction. Abduction is the inference type which tries to identify those initial conditions which deduction starts from, given general laws and some singular statements called *final states*. According to Lukasiewicz [56], abduction tries to "flow backwards along a deductive chain of inferences". Unfortunately, abduction is logically unsound since it involves a famous logical fallacy called "Affirming the Consequent". Since its logical unsoundness, we said that abduction "tries" to identify initial conditions because it is just able to "guess" these initial conditions and not to prove them. In the previous example, an abductive inference type could infer that "John is affected by acute hepatitis" starting from the general law "All patient affected by acute hepatitis manifest an increased value of serum transaminases" and from the singular statement "John manifests an increased value of serum transaminases", but nothing makes us sure of that, since other initial conditions may lead to an increased value of serum transaminases, such as miocardical infarction, or, in general, miopathies. Conclusions of an abduction are called "hypotheses".

Notice that the abduced statement "John is affected by acute hepatitis" may be viewed as a possible explanation for the statement "John manifests an increased value of serum transaminases". Otherwise, we have to stress that, following the Peirce's landmark definition, abduction is not collapsible on the "inference to the best explanation": "inference to the best explanation" is the inference process trying to identify the best explanation for a statement, while abduction, as an inference type, might be a building block of such an inference process, since it makes a guess about *possible* explanations.

Induction. Peirce distinguished three kinds of induction. The common feature of all kinds of induction is the ability to compare singular statements: induction is able to synthesize singular statements into general laws (i.e. infer from "John is affected by acute hepatitis" and "John manifests an increased value of serum transaminases" the general law stating that "All patient affected by acute hepatitis manifest an increased value of serum transaminases"), but it is also able to confirm or falsify (i.e. to verify) hypotheses on the basis of singular statements that can be deduced from them.

The inference model describes a plausible sequence of inference types applied to perform one of the above described basic tasks. Since we are interested in automatic reasoning, we do not claim that this inference model replicates the inference behavior of one or many human agents. From the standpoint of the epistemological foundation of KBS, an inference model may be defined as an abstract calculus acting on an ontology in order to execute the required task.

Medical reasoning, especially when executing the diagnostic task, has been convincingly modeled as a cycle of generating and testing hypotheses [57,6]. However, this definition does not allow us to understand how these hypotheses are generated and tested (i.e., it does not make clear which inference types are involved in such a schema and how they work). In the next section, we will argue that this inference model can execute the three main tasks involved in medical reasoning and will outline how this cycle may be interpreted as the result of chaining *abduction*, *deduction*, and *induction*.

fig2.eps3.5in

Figure 2: A schematic representation of the STMODEL.

3 An Epistemological Model for Medical Reasoning

This section describes a unified epistemological model of medical reasoning. This model is unified in the sense that the same inference model — we will call *Select* and *Test* Model (hereafter, STMODEL) — can execute the three generic tasks involved in medical reasoning. We will proceed as follows: first the general structure of the STMODEL will be described, and then we will illustrate how inference models involved in diagnosis, therapy planning and monitoring may be regarded as instances of the STMODEL.

3.1 The Select and Test Model

Advances in cognitive science have pointed out that the in a problem solving process experts usually select a small number of hypotheses and then they spent much time in testing and refining these hypotheses [58,9]. Medical reasoning may be broken down into two different phases: first, initial information is employed to select plausible hypotheses (*hypothesis selection phase*), and then these hypotheses are used as starting conditions to forecast expected consequences which should be matched with the state of affairs in the patient in order to confirm or falsify those hypotheses which they come from (*hypothesis testing phase*). Authors [59,29] agree about the utility of analizing this schema in terms of the inference types involved in it, as shown in Figure 2.

3.1.1 Abstraction

The process of selecting and testing hypotheses has been investigated by several AI researcher starting from de Groot's [60] pioneering studies on chess grand masters. A major result he achieved was that chess masters, before generating and testing the benefits of possible moves on the chessboard, undertake a preliminary recognition process in order to identify more abstract features of the board. From the standpoint of earlier cognitive psychology, this preliminary recognition process has been already conceptualized by Otto Selz [61] with the notion of *schematic anticipation*. In medical

14

problem solving, this phase corresponds to the preliminary identification the process of *abstracting* solution features from data. These interpretations are made with certainty: belief thresholds, and qualifying conditions are chosen so the abstraction is categorical in providing problem features.

3.1.2 Abduction

Starting from these problem features, abduction allows one to select plausible hypotheses. Two main kinds of abduction may be distinguished in this hypothesis generating phase:

1. *Unfocused Abduction*: This takes place at the beginning of the process, when general categories of hypotheses are selected to constrain the range of possible solutions. We refer to such an initial step as the formation of the initial hypothesis space.

2. *Focused Abduction:* This occurs when a hypothesis is selected in a context defined by another hypotheses. Such a step is qualified as either a refinement of the initial given hypothesis, or a competitor or a complement of that hypothesis.

Once hypotheses have been selected, they need to be ranked in order to speed up the testing phase, by testing first some preferred hypotheses. As already suggested by Peirce, multiple criteria may be taken into consideration for ordering hypotheses (i.e. preference criteria). In medical domain, these criteria could be: parsimony, danger, cost, curability, and so on. These criteria do not provide a way to select which hypotheses can be considered as solutions, but only a way to define the order in which they are tested.

3.1.3 Deduction

After hypotheses have been abduced and ranked, the testing phase starts to explore their consequences. Deduction allows us to derive from each candidate hypothesis what one expects to be true if that hypothesis is true. This kind of process is usually called *prediction*.

Deduction as logical consequence is the most conservative form of prediction: If a is true in some world W, then all logical consequences of a will be true in W. Many studies have argued that the classical interpretation of deduction as logical consequence is too weak to capture the aim of prediction [62,63]. For instance, one might conclude, from the facts that you are

stirring sugar into your hot tea and that sugar is soluble in your tea, that the sugar will dissolve in tea. But this conclusion holds only under certain other conditions, namely the condition "all other things being equal". The sugar would not dissolve if it were coated with wax or if there were already saturated solution in the tea or if other conditions occurred, some of which one could not even envision. Therefore conclusions we obtain are defeasible and hence they fail the property of monotonicity of logical consequence relation: once a conclusion has been drawn, it might have to be retracted when new information establishing that condition "all other things being equal" has been violated. Deduction we are dealing with is therefore a kind of nonmonotonic deduction, that is the basis of much work on nonmonotonic logics and belief revision [64].

3.1.4 Induction

Once predictions have been derived from hypotheses, they need to be matched in order to choose the *best* hypothesis. Induction is able to match single statement to single statement and, therefore, to match a single statement derived as a prediction from a hypothesis with a single statement describing a portion of the real state of affairs in the patient. Since hypotheses are ranked at the beginning of the testing phase, some hypotheses will be tested before others according to the adopted ranking criteria. During this phase, induction corroborates those hypotheses whose expected consequences turn out to be in agreement with the state of affairs in the patient and refutates those which failed this test. Induction closes the cycle of our inference model of medical reasoning.

The cyclic structure of the model stresses the nonmonotonic character of medical reasoning: previous hypotheses can be withdrawn whenever new information becomes available. As Stephanou and Sage [65] pointed out, uncertainty and imperfect information are fundamental characteristics of the knowledge relative to hypothetical reasoning. The nonmonotonic character of the STMODEL arises not only from the above mentioned nonmonotonic character of deductive inference type involved in it, but also from the logical unsoundness of the ascending part of the cycle guessing hypotheses to be tested. Doyle [66] pointed out that, since their unsoundness, these guesses do not exhibit the truth-preservative behavior of *ideal* rationality characterizing the incremental deduction of classical logic, but the nonmonotonic behavior of *limited* rationality of commonsense reasoning [67] that allows to discharge and abandon old hypotheses to make possible the tentative

/server/figures/diagnosis.eps3.5in

Figure 3: Specializing the STMODELin the diagnostic task.

adoption of new ones. Notice that this adoption is not merely tentative but rationally tentative, in the sense that, just as abduction, it is based on a reasoned selection of knowledge [68] and on some preference criteria which avoid the computational explosion of hypotheses generation. One of the principal means of limiting rationality is indeed to limit efforts by directing attention to some areas and ignoring others. This character matches exactly with the ability of an expert in generating a small set of hypotheses to be carefully tested. But in such a case, the expert has to be ready to withdraw paths of reasoning when they diverge from the correct path, that is from the path that would have taken the expert had considering the ignored knowledge portions). In such a way, the nonmonotonic character turns out as a foundative epistemological feature of the STMODEL of medical reasoning, since this nonmonotonic character is the result not of a mere lack of information but of a reasoned limiting of information imposed by its own logical unsoundness.

3.2 Tasks in Medical Reasoning

Following a slightly modified version of the Chandrasekaran's definition, each basic task executed in medical reasoning (i.e diagnosis, therapy planning and monitoring) requires the combination of an ontology and an inference model. Diagnosis, therapy planning and monitoring are therefore executed by an inference model representing an instance of the STMODEL working on a specific ontology.

3.2.1 Diagnosis

Figure 3 shows the way of specializing STMODEL for the diagnostic task. Diagnostic reasoning provides an explanation of the current situation in the patient. Therefore, it usually represents the first task to be executed in medical reasoning, since therapy planning and monitoring need a previous understanding of the situation before their execution.

Diagnosis starts from *observed data* that usually provide incomplete information about the state of affairs in the patient. This information is subsequently mapped into *clinical evidences to be explained*. Then, abduc-

fig4.eps3.5in

Figure 4: Specializing the STMODELin therapeutic task.

tion selects plausible *diagnostic hypotheses*. In diagnostic task, abduction is said to be unfocused when it takes place in an empty diagnostic space and evokes categories of hypotheses in order to constrain the range of possible diagnoses for the case. A focused abduction occurs when hypotheses are selected in the context defined by another diagnostic hypothesis. After abduction has taken place, the set of diagnostic hypotheses (i.e., the *diagnostic space*) needs to be organized for planning the next step of reasoning.

Starting from highest ranked diagnostic hypothesis, deduction decides which findings have to be expected if this hypothesis is true (*expected data*). Thus, new laboratory or clinical examinations can be requested to verify unobserved expectations. Finally, induction establishes whether hypotheses can be concluded or refutated, or whether they are worth testing further, depending on how closely the observed findings match expectations. Furthermore, this inference type deals with the termination of the diagnostic process: it decides whether a *satisfactory* explanation of the patient's state has been achieved.

3.2.2 Therapy Planning

Therapy planning may be represented as shown in Figure 4. As a planning problem, it may be ragarded the selection of the best set of actions (i.e. therapies) leading to a future goal (i.e. ameliorate the condition in the patient). It starts with the *diagnosis* and the *observed data* available on the patient at hand. From these information, a process of abstraction derives a concise and essential portrait of the situation consisting of a list of *therapeutic problems* which may be relevant for the given patient. Such a task not only involves mapping continuous values of clinical variables into meaningful categorical propositions, but also, and more important, deriving a restricted set of critical aspects of the patient's condition which can be immediately interpreted as a list of crucial targets of the therapy.

The successive abduction takes the list of therapeutic problems and infers a presumptive list of *therapies* which includes those treatments that deserve consideration as potentially useful in handling those problems. Far from being definitive, elements of this list are considered just as hypotheses, needing further more focused analysis and testing. In fact, reasoning proceeds with

a ranking phase, which usually establishes priorities over the current list of treatments. The testing phase involves the deductive-inductive inference. Deduction here consists in focusing on single treatments of the list, or pairs of them, in order to perform a more thorough evaluation of their appropriateness for the patient at hand. This usually involves making predictions in order to estimate possible consequences of the treatments on the clinical course of the specific patient. As stressed above, this requires nonmonotonic reasoning (i.e. involves the assumption "being all things equal"). New data usually trigger further cycles, until the list of treatments is reduced to the point of providing a helpful advice.

3.2.3 Monitoring

Although the word monitoring is often thought to refer to the action of obtaining data, the proper use of the term in medicine is to observe and to control the course of a patient's condition. It is essential to answer here the following question: does monitoring represent a different generic task from diagnosis and therapy planning? Without taking into consideration very specific situations, diagnosis can be considered as the task of achieving the best explanation of patient's condition, therapy planning the best action to perform in order to improve patient's condition, and monitoring the best strategy to verify if the planned action proves to be really effective. To this aim, a KBS should be able to predict the course of a patient's condition under the combined action of diagnosed disorders and selected therapy. This ability should also characterize the action of an expert.

From an epistemological point of view, monitoring may be described by the STModel. If the selected therapy works and the patient is responding appropriately, according to the specific patient model used, then therapy is continued or the patient is released from treatment (*testing phase*). If the therapy did not work or if unusual findings arise, then further assessment is necessary (*selecting phase*). As a result of monitoring, previous diagnosis and therapy planning tend to be either confirmed or rendered questionable. In the former case monitoring implies continuous cycling between deduction and induction, while in the latter case diagnosis and/or therapy planning may need to be executed again, so requiring abductive inferences starting from the new patient's condition.

4 The Computational Level

At the computational level, the major problem of medical reasoning is to choose methods and formalisms for implementing a given epistemological model. The aim behind this section is twofold. First, we wish to argue that the STMODEL may improve the productivity of KBS development and knowledge acquisition by making clear which aspect of medical reasoning are intrinsic to the generic tasks to be executed and which artifacts of the implementation are imposed by chosen methods and knowledge representation formalisms. As a matter of fact, the STMODEL permits the knowledge engineer and the expert to cooperatively develop KBS using a shared language of epistemological entities rather than terms of the underlying implementation. Second, we argue that "many methods are better than one", that is, transforming a formalism in a general purpose tool limits the expression of features and components of the STMODEL so severely that this tool is no longer applicable or, at least, desirable.

Here we will briefly describe basic computational features of the architectures of some well-known systems in terms of STMODEL to point out aspects which seem to be intrinsic to the underlying epistemological model and not exclusively related to the problem they try to solve.

4.1 Hypothesis Selection

Problem. Which methods are available for generating hypotheses from observed findings or manifestations? Which are their strengths and serious limitations?

Discussion. Different approaches for generating hypotheses which employ abduction have been proposed in the literature. We will consider the most interesting ones from the perspective of available methods for implementing the generating phase of the STMODEL, that is the ascending pathway leading from patient's data to the set of abduced hypotheses.

Current research on hypothesis generation emphasizes *set covering* approaches [69,70,71]. Although different authors call the hypotheses space differently - "hypotheses graph", "candidate graph", and "hitting set tree" - their algorithms are categorical and share the same underlying structure. They are pure syntactic methods generating candidates which form the subset of disorders, each representing a possible explanation for a given set of manifestations.

The problem here is to deal with multiple disorders, that is with candidate assembly, since finding all minimal set covers by assembling hypotheses is NP-hard [72]. An additional problem arises from the fact that the process of candidate generation sometimes depends on the order in which manifestations are entered: some ordering causes an oscillatory phenomenon of generating and pruning at each frontier and a computational barrier from one frontier to the next When set covering models are used in domains with complex ontologies, typically in medicine, difficulties in dealing with multiple disorders suggest to use probabilistic methods to transform categorical algorithms into heuristic but tractable ones [73,70,74].

In recent years, great efforts have been devoted to the development of a probabilistic scheme where the structure of domain knowledge is represented through a directed graph, such as Bayes Belief Network (BBN) [75].

The key computational feature of BBN is the fact that they allow us to reduce the computation to a series of local calculations using only variables obtainable from one entity and its neighbors in a graph structure. Hereby a calculation of the global joint probability distribution is avoided. When the network is singly connected (i.e., at most one path between pairs of nodes is allowed) Kim and Pearl [76] developed an algorithm for propagating uncertainty in a BBN. Unfortunately, finding the most probable hypothesis in BBN has been shown to be intractable in general [77]. Therefore, few real medical problems can be solved through singly connected BBN, and multiply connected networks (i.e., more than one path may exist between a pair of nodes) are required to represent the domain knowledge.

The development of MUNIN [78,79], a KBS for electromyography, required the construction of networks containing several hundred nodes. Handling such large networks raised new computational problems needing efficient solutions. Shwe and Cooper [80] are presently developing a BBN-based reformulation of the knowledge base of Quick Medical Reference (QMR) [81], a KBS for diagnosis in internal medicine that was developed at the University of Pittsburgh as the successor to INTERNIST-1 [4]. To deal with the representational and computational complexity of QMR-DT, several simplifying assumptions have been made, such as marginal independence of diseases, conditional independence of findings given any hypothesis of disease, and the assumptions that findings are manifestations of disease.

The approaches described above share common problems. They do not allow us to exploit a preliminary data abstraction, and then use a focusing strategy to reduce the set of hypotheses to be considered. They behave like novice clinicians that are inefficient clinical problem solvers even though a

simple and clear ontology is available. They try to solve the problem by evoking all possible disorders and then clustering them.

4.2   Hypothesis Testing

Problem. Is modeling the course of a patient's condition, under the action of pathogenetic mechanisms and therapies, essential for hypothesis testing?

Discussion.   As argued in this paper, hypothesis testing represents an essential part of medical reasoning and involves deductive-inductive inferences. There are numerous competing modeling approaches that attempt to express available medical knowledge for predicting what is expected to be observed in a patient under a hypothesis derived from a previous abductive inference.

Computational models may be described in terms of expressiveness and computational tractability. Model expressiveness refers to the breadth of concepts that can be represented using that formalism. Computational tractability refers to the ease with which predictions can be derived from internal model structures. A modeling methodology that is expressively rich but computationally intractable is not useful in constructing real KBS. Model expressiveness and computational tractability tend to be opposing characteristics of a modeling method. Thus, the developer of a KBS must be able to choose the most suitable method by making a difficult tradeoff taking into consideration goal, inference model and ontology of the generic task he/she wants the KBS to be able to execute.

Modelling world requires to place a special emphasis on the concept of time. There two opposing perspectives about the nature of time that philosophers, logicians, and AI researchers have used to justify particular models of time. One perspective assumes that time is a real entity with objects and properties that are unique to it. This perspective leads to models where time itself is the essential element to be characterized and manipulated. The opposite perspective argues that time exists only because the existence of change; therefore, a model of change is a model of time. This second approach yields models that express temporal concepts in terms of processes and the changes these processes cause. Time is an artifact or a side effect rather the primary object. The distinction is important because the choice of perspective has a major influence on the terms, relations, and operations that are provided within a model representation language.

Formalisms modeling time as the primitive concept differ in the unit of time that is considered to be the basic temporal entity: instantaneous points or moments [82,83] versus intervals or durations [84,85]. The choice of temporal points or intervals is an ontological question. Few attempts have been made until now to use in medical KBS an interval-based representation to construct a model of the key events that occur during a patient's clinical course [86,87]. Modeling methods based on a representation of change can be assigned to two different classes: models relating states or events to each other and models providing predictions of pathophysiological system behavior based on a representation of its structure. Formalisms like associative causal networks and BBN belong to the first class, however their ability to manage time raised difficult problems [88]. Models based on quantitative and qualitative differential equations fall in the second class.

The ontological difference between the two modelling approaches, that is, modelling time-intervals versus modelling change, reflects different goals. The first type of models attempts to capture the temporal relationship between facts in the world; the others wish to capture the temporal effects of pathophysiological mechanisms or therapeutic actions on the different components of a system. In medical reasoning both perspectives make sense. The patient's medical history consists of facts that change over time as the patient's illness evolves and responds to administered therapy. This perspective fits well with state- or event-based models. Conversely, disease manifestations and therapy management are the result of pathophysiological, homeostatic, and therapeutic processes operating simultaneously with synergistic and antagonistic interactions. Reasoning at this level fits well with mechanistic models of pathophysiology as interaction of processes. We believe that there is need for both perspectives and that any KBS that cannot represent both views of time is likely to be deficient in its reasoning capabilities.

For representing and exploiting pathophysiological models in diagnostic KBS, a qualitative formulation has to be preferred to a quantititive one because of the reduced number of assumptions in building the model and because of its ability to reason from pathophysiological principles. Moreover, traditional quantitative models provide the numeric values that the model variables will assume over time, and these provide the predicted behavior of a real system only if model parameters can be accurately estimated from experimental data. This situation occurs very rarely in medicine.

Several approaches to the formulation and simulation of qualitative models have been proposed and implemented. QSIM has been the widest ap-

plied in medical domain, to model physiological systems. QSIM provides a descriptive language, consisting of qualitative constraints that abstract the relationships in a differential equation, to represent the structure of a pathophysiological system and a simulation algorithm to infer its qualitative behaviour. Qualitative predictions are obtained from the model by perturbing the normal state according to the hypothesized pathogenetic action of the disease. Fault models corresponding to multiple diagnostic hypotheses are easy to build since it is possible to submit a model to any number of perturbations. When a disease is hypothesized, one or several fault models are taken into consideration: each model provides an aetiological theory of the observed patient's state. The perturbations to the normal state define the initial state of the simulation.

Therapy planning often involves complex, conflicting objectives. For any decision involving conflicting objectives, the optimal decision represents a tradeoff among the competing goals. A KBS, if it has to make recommendations for actions, must be able to access and compare the costs and benefit of the possible effects of a therapy by taking into consideration uncertainty about the patient's course given that therapy and utilities associated to the therapy itself and its possible consequences. Decision theoretic models [89] seem appropriate in these cases. Their advantages are that patient utilities, incorporating preferences and risk attitudes, are modeled explicitly, and that probabilistic representation of uncertainty is axiomatic and provides the range of quantification necessary to represent the small but important differences in likelihood of the different effects of administered therapy.

4.3 Many Methods are Better than One

Problem. Is there a unique method able to meet all requirements, both representational and computational, needed to develop a KBS?

Discussion. Computational issues discussed in previous sections support the argument that there is not a unique knowledge formalism which allows us to implement a KBS able to develop medical reasoning as well as has been described at the epistemological level. A very promising approach based on frame-like languages has been proved to be too weak both in expressivity and in computational tractability. The failure of the recent attempt to represent ABEL's causal ontology using these languages [90] proved their expressivity limits. Furthermore, other experiments using the same languages showed that, in most critical applications, they produce constructs requiring

24

fig5.eps3.5in

Figure 5: The blackboard architecture of NEOANEMIA.

non-polynomial (or otherwise unacceptably long) worst-case response times
for sound and complete classification of concepts in taxonomic hierarchies.
Levesque and Brachman [91] suggest to omit these constructs. However,
Doyle and Patil [64] convincingly demonstrated that these restrictions de-
stroy the generality of the language. Definitions inexpressible because of
language restrictions must be entered as primitive concepts, which are un-
classifiable, and thus reduce the utility of classification. Moreover, restrict-
ing classification to purely definitional information significantly reduces its
utility in domain applications. STMODEL allows us to claim that the solu-
tion for this problem stays beyond formalisms and it must be sought in the
epistemological level.

Ontological and computational issues force us to combine different for-
malisms and methods in KBS implementation. Namely, the main com-
putational problem seems to be the way these different methods can be
combined to deal with different ontological features and different phases of
reasoning. The aim, here, should be the development of tools allowing the
implementation of an architecture embedding different knowledge represen-
tation formalisms and a separate control module, able to choose dynamically
among those formalisms according to the ontology of the domain, the phase
of reasoning to be performed, and the available computational resources.

The blackboard control architecture provides a suitable framework for
building KBS fulfilling the requirements mentioned above [92]. It extends
and elaborates the standard blackboard architecture [93] in order to ap-
proach the problem of controlling the KBS's problem solving process as a
real-time planning problem.

Briefly, the blackboard control architecture extends and elaborates the
architectural features summarized above as follows. It defines explicit do-
main and control blackboards. Moreover, it defines explicit domain and con-
trol knowledge sources which operate primarily on the domain and control
blackboard, respectively. This allows us to represent explicitly and sepa-
rately the inference model (i.e. control knowledge) and the ontology (i.e.
domain knowledge) and to choose the most suited representation formalism
for each knowledge source.

Figure 5 shows the architecture of NEOANEMIA [54], a KBS able to
recognize disorders causing anemia. It derives from a previous KBS called

25

ANEMIA [94] which was proved to perform quite satisfactorily [95]. The control blackboard contains information about which strategies have been applied to obtain the present solution of the diagnostic problem thus allowing the system to be aware of the problem solving strategy it is exploiting and to switch to another one if the previous one did not prove to be effective, while the domain blackboard contains the solution elements organized at different level of abstraction. According to the epistemological model of the diagnostic task, NEOANEMIA starts selecting, via unfocused abduction, general diagnostic categories from clinical evidences obtained through an abstraction of routine hematologic findings. All NEOANEMIA's entities, i.e. patient's data and diagnostic entities, are organized into taxonomic structures. The system proceeds by exploring (*unfocused abduction*) and possibly refining among the previously abduced hypotheses (*focused abduction*). During its abductive inference, NEOANEMIA exploits a knowledge base built up using production rules and generates a diagnostic space containing all the abduced diagnostic hypotheses. Then, diagnostic space structuring takes place: NEOANEMIA mainly looks for compatibilities or possible associations, thus grouping or differentiating among the diseases included in the diagnostic space. In such a way, composite hypotheses can be assembled.

Once abduction and diagnostic space structuring have been executed, NEOANEMIA will deduce expected manifestations from the hypotheses included into the diagnostic space. While in the hypothesis generation phase, NEOANEMIA exploits compiled heuristic pathways specified by the expert, a separate and explicit representation of taxonomic and causal ontology is used in the testing phase. They have been represented using respectively a simple two layer network (i.e. representing which clinical evidences may be expected in presence of each disease) and QSIM [50] for representing available knowledge on pathophysiological system dynamics. QSIM requires that the basic aetiological mechanisms of a disease are represented as perturbations of the initial conditions of a qualitative model of iron metabolism and expected manifestations are then found by analysing the new reached steady state conditions. QSIM models provide a deeper understanding of portions of knowledge expressed by the two level network. They also predict the effect of multiple diseases acting together.

The inductive phase is then carried out by comparing expected with observed findings, when available, or by planning what to do next. Further actions aim either to corroborate candidate hypotheses or to discriminate between competing hypotheses, or to rule-out hypotheses. The selected action depends on the contents of the structured diagnostic space and the

strategy to make the decision is represented in the control knowledge by meta-rules. In order to represent explicitly the taxonomic ontology exploited in the deductive phase, we are enhancing the frame-based language we used with an ATMS [96], allowing us to handle the nonmonotonic character of the deductive phase: the hypothesis hierarchy is mapped into a hierarchy of environments each representing different possible states of the patient as suggested by abduced hypotheses.

Computational problems arising in the implementation of the epistemological model of therapy planning are presently under careful consideration. To this aim a KBS called Therapy Advisor (TA) has been developed [97]. The task of TA is planning an adequate therapy depending on the etiopathologic mechanisms causing a patient's anemia as diagnosed by NEOANEMIA.

According to the epistemological model of the therapy planning task, TA starts from deriving a restricted set of critical aspects of a patient's condition which might be eliminated adopting a suitable therapy. These represent abstractions which have been cathegorically derived from patient's data and formulated diagnosis in order to abduce a set of therapies potentially useful to handle those problems. For most of them a well agreed-on therapeutic plan can be proposed. In such cases, therapy planning does not involve basic strategic choices, but rather consists in fixing details such as dosage and route of administration in order to meet patient's conditions. Thus, production rules have been judged as the most appropriate formalism for representing knowledge. The remaining therapeutic problems involve trade-offs between conflicting goals, in presence of uncertainty about the therapy effects.

In presence of uncertainty and trade-off, our conclusion about the recommended therapeutic action may depend on a large number of patient's data. They are usually called "defeaters". For example, the decision to remove the spleen (splenectomy) in a myelofibrotic patient in order to reduce his/her transfusional need interacts with the "platelet count", since in presence of a high platelet count there is a significant risk of splenectomy-induced thrombosis.

The presence of several defeaters would burden the rules with a large number of possible exceptions whose pathophysiological meaning would be difficult to be understood by a non-expert user. Moreover, it is essential for building an advanced TA to represent explicitly and separately the medical knowledge the system uses to predict the effects of a therapy, possibly through a model of the underlying pathophysiological system, and the preferences, elicited from either the user or the patient.

These reasons led to the choice of adopting "influence diagrams" (IDs) for representing the most challenging therapeutic decision problems. IDs are directed acyclic graphs with three types of nodes: decision, chance, and utility nodes. Decision nodes represent choices available to the decision-maker. Chance nodes represent uncertain quantities, that may either characterize the patient's condition at the decision time or therapy outcomes. Finally, the utility node embeds the utilities assigned to possible configurations of values of chance nodes.

TA's knowledge base contains some taxonomies of entities represented by means of frames: patient's findings and diagnostic entities are shared with NEOANEMIA, while therapeutic problems and adoptable therapies are specific portions of knowledge for therapy planning. Starting from observed patient's data and diagnosis formulated by NEOANEMIA, TA selects the relevant therapeutic problems in the patient at hand. To this aim TA exploits different kinds of pathophysiological concepts: some of these are abstracted by NEOANEMIA from patient's data (for example, "severe anemia", "augmented serum bilirubin", etc.), some others represent general categories of diseases included by NEOANEMIA into its diagnostic space (for example, "aregenerative anemias", "hemolytic anemias", etc.) able to define in a synthetic way the behavior of the erythropoietic system, while the remaining concepts are abstracted by TA from patient's data and are needed for selecting the most suitable therapy.

Two fundamental design choices have been made from the computational point of view. First, TA was built exploiting the blackboard control architecture framework, as well as we did developing NEOANEMIA. Second, IDs have been used for representing a set of significant therapeutic decision problems. This limited the size of each ID, so avoiding the need to tackle any problem of efficiency in managing uncertainty on a large BBN. Onthe other hand, TA was designed so that it can provide support to retrieve the ID that is appropriate to the particular decision at hand.

Investigating architectural foundations for real-time performance in intelligent agents, Hayes-Roth [98] proposed an architecture for these agents comprising subsystems for perception, cognition and action. Focusing our attention only on the cognition subsystem, we can say that it holds all of an agent's knowledge and performs all of its reasoning. It asynchronously acquires perceived information, retrieved from its input buffers, and performs a variety of knowledge-based reasoning steps, among which the following are interesting for us: detection and interpretation of events (*diagnosis*), planning long-term courses of action (*therapy planning*), and adaptation of its

behavior to changing environmental conditions (*monitoring*).

GUARDIAN [99], developed on top of BB1, performs all of these three generic tasks. For some of these tasks, it uses both compiled and explicit ontology and inference model. The compiled pathways permit quick responses to familiar situations. The explicit ontology embeds more fundamental biomedical knowledge and permits more thorough (and time-consuming) responses to both familiar and unfamiliar cases. Compiled and explicit knowledge are implemented as sets of abstract reasoning operations that are triggered by particular kinds of perceptual or cognitive events, along with control operations that construct resource-bounded control plans in particular context.

5   Conclusions

This paper presents a new abstraction paradigm aiming at unifying different ideas that have been recently proposed for the analysis and design of KBS. These ideas all attempt to provide descriptions of KBS reasoning for solving a particular kind of problems at a conceptual level that is above the implementation, thus making clear which aspects of a class of problems are intrinsic to the problem and which are artifacts of the implementation. Four fundamental ideas have been carefully considered: inference structures, deep/shallow system, problem solving method, and generic tasks. Each of these ideas focuses on a particular feature of reasoning: the pattern of inferences, the domain models underlying expertise, the sequence of actions needed to solve a problem, and the task features. Our proposal is based on a two levels analysis of KBS: an epistemological and a computational level.

At the first level, ontology and inference model of a KBS have to be defined. Ontology represent the conceptual model of entities and relationships composing the domain knowledge, while the inference model is the conceptual representation of the inference structure to execute a task by managing that ontology. At the computational level, methods and formalisms selected to build and exploit a knowledge base fall. We outlined the fact that different disciplines (i.e., AI, logic, probability theory, decision theory, mathematics, etc.) can provide suitable formalisms and methods. They should be adopted after the epistemological analysis has been carried out and taking into account the constraints which derive from the size and the type of available domain knowledge and from the environment where the KBS should operate (i.e., small time intervals allowed to make decisions and limited computer

resources).

We constrained our study to medicine and identified three generic tasks: diagnosis, therapy planning and monitoring. They are obviously interrelated but can be submitted to a separate epistemological analysis. The main result of this analysis was that these generic tasks manage different ontologies, but they can be executed exploiting a unique inference model. Such a model involves three different inference types: abduction, deduction, and induction. Thus, medical reasoning may be broken down into two different phases: first, initial information is used to generate (abduce) plausible hypotheses (hypothesis selection phase), then these hypotheses are used as starting conditions to forecast (deduce) expected consequences, then these expected consequences are matched with the state of affairs in the patient in order to confirm or falsify (induction) those hypotheses which they come from (hypothesis testing phase).

References

[1] S. G. Pauker, G. A. Gorry, J. P. Kassirer, and W. B. Shwartz, "Toward the simulation of clinical cognition: Taking a present illness by computer," *The American Journal of Medicine*, vol. 60, pp. 981–995, 1976.

[2] M. C. Weiss, C. A. Kulikowsky, and A. K. Safir, "Glaucoma consultation by computers," *Computers in Biology and Medicine*, vol. 1, pp. 25–40, 1978.

[3] E. Shortliffe, *Computer Based Medical Consultations: MYCIN*. Amsterdam: North Holland, 1976.

[4] R. A. Miller, H. E. Pople, and J. D. Myers, "Internist-1. an experimental computer based diagnostic consultant for general internal medicine," *New England Journal of Medicine*, vol. 307, pp. 468–476, 1982.

[5] R. Davis, "Diagnostic reasoning based on structure and behavior," *Artificial Intelligence*, vol. 24, pp. 347–410, 1984.

[6] H. Simon, "Artificial-intelligence approaches to problem solving and clinical diagnosis," in *Logic of Discovery and Diagnosis in Medicine* (K. F. Shaffner, ed.), pp. 72–93, Berkely and Los Angeles: University of California Press, 1985.

[7] M. Genesereth, "The use of design descriptions in automated diagnosis," *Artificial Intelligence*, vol. 24, pp. 411–436, 1984.

[8] D. B. Lenat, M. Prakash, and M. Shepherd, "Cyc: Using commonsense knowledge to overcome brittleness and knowledge acquisition bottlenecks," *AI Magazine*, vol. 6, pp. 65–85, 1986.

[9] K. VanLehn, "Problem solving and cognitive skill acquisition," in *Foundations in Cognitive Science* (M. I. Posner, ed.), pp. 527–579, Cambridge (MA): MIT Press, 1989.

[10] V. L. Patel, G. J. Groen, and M. Ramoni, "Cognitive depth versus psycological depth," *Artificial Intelligence in Medicine*, To appear.

[11] E. T. Keravnou and J. Washbrook, "Deep and shallow model in medical expert systems," *Aritificial Intelligence in Medicine*, vol. 1, pp. 11–28, 1989.

[12] D. Karp and Wilkins, "An analysis of the distinction between deep and shallow expert systems," tech. rep., STAN-KSL-89-10. Knowledge Systems Laboratory. Stanford University. Stanford (CA), 1989.

[13] D. Forsythe and B. G. Buchanan, "Knowledge acquitision for expert systems: Some pitfalls and suggesions," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, pp. 435–442, 1989.

[14] T. Gruber and P. Cohen, "Design for acquisition: principles of knowledge-system design to facilitate knowledge acquisition," *International Journal on Man-Machine Studies*, vol. 26, pp. 143–159, 1987.

[15] L. Steels, "Components of expertise," *AI Magazine*, vol. 12, pp. 38–49, 1990.

[16] A. Newell, "The knowledge level," *Artificial Intelligence*, vol. 18, no. 1, pp. 87–127, 1982.

[17] W. J. Clancey, "Heuristic classification," *Artificial Intelligence*, vol. 27, pp. 289–350, 1985.

[18] B. Chandrasekaran and S. Mittal, "Deep versus compiled knowledge in diagnostic problem solving," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 349–354, Morgan Kauffman, 1982.

[19] J. McDermott, "Preliminary steps toward a taxonomy of problem solving methods," in *Automating Knowledge Acquisition for Expert Systems* (S. Marcus, ed.), pp. 225–256, Boston (MA): Kluwer, 1988.

[20] S. Marcus, "Salt: A knowledge-acquistion tool for pourpose-and-revise systems," in *Automating Knowledge Acquisition for Expert Systems* (S. Marcus, ed.), pp. 81–123, Boston (MA): Kluwer, 1988.

[21] B. Chandrasekaran, "Towards a taxonomy of problem solving types," *AI Magazine*, vol. 4, pp. 9–17, 1983.

[22] W. J. Clancey, "NEOMYCIN: Reconfiguring a rule-bases expert system from application to teaching," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 829–836, 1981.

[23] W. Clancey, "Extensions to rules from explanation and tutoring," in *Rule-based Expert Systems*, pp. 351–371, Readings (MA): Addison-Wesley, 1984.

[24] H. Levesque, "Knowledge representation and reasoning," *Annual Review on Computer Science*, vol. 1, pp. 255–287, 1986.

[25] L. Eshelman, "MOLE: A knowledge-acquisition tool for cover-and-differentiate systems," in *Automating Knowledge Acquisition for Expert Systems* (S. Marcus, ed.), pp. 37–80, The Netherlands: Kluwer Academic Publishers, 1989.

[26] B. Chandrasekaran and S. Mittal, "Conceptual representation of medical knowledge for diagnosis by computer: Mdx and related systems," in *Advances in computers* (M. Yovits, ed.), pp. 217–293, New York: Academic Press, 1983.

[27] B. Chandrasekaran, "Generic tasks in knowledge-based reasoning: high," *IEEE-Expert*, vol. 1, pp. 23–30, 1986.

[28] B. Chandrasekaran, "Generic tasks as building blocks for knowledge-based systems: The diagnosis and routine design examples," *The Knowledge Engineering Review*, vol. 3, no. 3, pp. 183–210, 1988.

[29] L. Johnson and E. T. Keravnou, *Expert Systems Architecture*. Koogan Page: 1988, London.

[30] S. Norrel, "Models of causation in epidemiology," in *Health, Disease and Causal Explanations in Medicine* (L. N. et al, ed.), pp. 147–152, Dordrecht: Reidel, 1984.

[31] R.-Z. Qui, "Models of explanation and explanation in medicine," *International Studies in the Philosophy of Science*, vol. 3, pp. 199–212, 1989.

[32] P. Ridderikhoff, *Methods in Medicine. A descriptive study of Physicians' Behavior*. Kluwer Academic Publisher: 1989, Dordrecht.

[33] J. L. Kolodner and R. M. Kolodner, "Using experience in clinical problem solving: Introduction and framework," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, pp. 420–431, 1987.

[34] V. L. Patel, G. J. Groen, and Y. Patel, "Cognitive aspects of clinical performance: The role of clinical expertise," *American Journal of Medicine*, To appear.

[35] F. Strawson, *Individuals: An Essay of Descriptive Methaphysics*. London: Methuen, 1959.

[36] C. Glymour, "When less is more," in *Cognitive Science in Medicine. Biomedical Modelling* (D. A. Evans and V. L. Patel, eds.), pp. 349–367, Cambridge (MA): MIT Press, 1989.

[37] P. J. Hayes, "Naive physics i: ontology for liquids," in *Formal Theories of Commonsense World* (J. R. Hobbs and R. C. Moore, eds.), pp. 89–93, Norwood (NJ): Ablex, 1985.

[38] J. McCarthy, "Circumscription — a form of nonmontonic reasoning," *Artificial Intelligence*, vol. 13, pp. 27–39, 1980.

[39] J. Alexander, M. J. Freling, S. J. Shulman, J. L. Staley, S. Rehfuss, and S. L. Messick, "Knowledge level engineering: ontological analysis," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 963–968, Morgan Kauffman, 1986.

[40] D. B. Lenat, "Ontological versus knowledge engineering," *IEEE Trans. Knowledge and Data Engeneering*, vol. 1, pp. 84–88, 1989.

[41] J. McCarthy and P. J. Hayes, "Some philosophical problems from the standapoint of artificial intelligence," in *Machine Intelligence 4*

(B. Meltzer and D. Mitchie, eds.), pp. 463–502, Edinburgh: Edinburgh University Press, 1969.

[42] R. Milne, "Strategies for diagnosis," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, pp. 333–339, 1987.

[43] B. Chandrasekaran, J. W. Smith, and J. Sticklen, "'deep' models and their relation to diagnosis," *Artificial Intelligence in Medicine*, vol. 1, pp. 29–40, 1989.

[44] B. J. Kuipers, "Qualitative simulation as causal explanation," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, pp. 432–444, 1987.

[45] H. E. Pople, "Evolution of an expert system: From INTERNIST to CA-DUCEUS," in *Proceedings of the International Conference on Artificial Intelligence in Medicine* (I. D. Lotto and M. Stefanelli, eds.), (Amsterdam), pp. 128–208, North Holland, 1985.

[46] R. Patil, P. Szolovits, and W. B. Schwartz, "Causal understanding of patient illness in medical diagnosis," in *Proceedings of the International Joint Conference on Artificial Intelligence*, (Vancouver, British Columbia, Canada), pp. 893–899, Aug. 1981.

[47] B. J. Kuipers, "Qualitative simulation," *Artificial Intelligence*, vol. 29, pp. 280–338, 1986.

[48] J. de Kleer and J. Brown, "A qualitative physics based on confluences," *Artificial Intelligence*, vol. 24, pp. 7–83, 1983.

[49] K. Forbus, "Qualitative process theory," *Artificial Intelligence*, vol. 24, pp. 85–168, 1984.

[50] L. Ironi, M. Stefanelli, and G. Lanzola, "Qualitative models in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 2, pp. 85–101, 1990.

[51] E. Rosch and C. B. Mervis, "Family resemblance: Study in internal structure of categories," *Cognitive Science*, vol. 7, pp. 573–605, 1975.

[52] J. S. Aikins, "Prototypical knowldge for expert systems," *Artificial Intelligence*, vol. 20, pp. 163–210, 1983.

[53] K. F. Shaffner, "Introduction," in *Logic of Discovery and Diagnosis in Medicine* (K. F. Shaffner, ed.), pp. 1–36, Berkely and Los Angeles: University of California Press, 1985.

[54] G. Lanzola, M. Stefanelli, G. Barosi, and L. Magnani, "Neoanemia:a knowledge-based system emulating diagnostic reasoning," *Computers and Biomedical Research*, vol. 23, pp. 560–582, 1990.

[55] C. S. Peirce, *Collected Papers*, vol. 2. Cambridge (MA): Harward University Press, 1931.

[56] J. Lukasiewicz, "Creative elements," in *Selected Works*, Amsterdam: North Holland, 1970.

[57] A. Elstein, L. Shulman, and S. Sprafka, *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge (MA): Harvard University Press, 1978.

[58] M. Chi, R. Glaser, and M. Farr, *The Nature of Expertise*. Hillsdale (NJ): Erlbaum, 1988.

[59] J. F. Sowa, *Conceptual Structures. Information Processing in Mind and Machine*. Reading (MA): Addison-Wesley, 1984.

[60] A. deGroot, *Though and choice in chess*. Mouton: The Hague, 1978.

[61] O. Selz, "The revision of the fundamental conceptions of intellectual processes (1927)," in *Thinking: From Association to Gestalt*, New York: Wiley, 1968.

[62] G. Harman, *Change in View. Principles of Reasoning*. Cambridge (MA): MIT Press, 1986.

[63] D. Poole, "Explanation and prediction: an architecture for default and abductive reasoning," *Computational Intelligence*, vol. 5, pp. 97–110, 1989.

[64] J. Doyle, "Constructive belief and rational representation," *Computational Intelligence*, vol. 5, pp. 1–11, 1989.

[65] H. Stephanou and A. Sage, "Perspectives in imperfect information processing," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, pp. 780–798, 1987.

[66] J. Doyle, "Artificial intelligence and rational self-government," tech. rep., CMU-CS-88-124, Computer Science Department, Carnegie Mellon University, 1988.

[67] H. Simon, *The Science of Artificial.* Cambridge (MA): MIT Press, 1969.

[68] C. Truesdell, *An Idiot Fugitive Essay on Science: Criticism, Training, Circumstances.* Berlin: Springer-Verlag, 1984.

[69] J. A. Reggia, D. S. Nau, and P. Y. Wang, "Diagnostic expert systems based on set covering model," *International Journal of Man-Machine Studies*, vol. 19, pp. 437–460, 1983.

[70] J. de Kleer and B. Williams, "Diagnosing multiple faults," *Artificial Intelligence*, vol. 32, pp. 97–130, 1987.

[71] R. Reiter, "A theory of diagnosis from first principles," *Artificial Intelligence*, vol. 32, pp. 57–95, 1987.

[72] T. Bylander and J. Smith, "The computational complexity of abduction," *Artificial Intelligence*, 1991.

[73] Y. Peng and J. A. Reggia, "A probablistic causal model for diagnostic problem solving i: Integrating symbolic causal inference with numeric probabilistic inference," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, pp. 146–162, 1987.

[74] T. D. Wu, "Symptoms clustering and syndrome knowledge in diagnostic problem solving," *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, 1989, pp. 45-49.

[75] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artificial Intelligence*, vol. 29, pp. 241–288, 1986.

[76] J. Kim and J. Pearl, "A computational model for causal and diagnostic reasoning in inference systems," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 280–338, 1983.

[77] G. Cooper, "The computational complexity of probabilistic inference using bayesian belief networks," *Artificial Intelligence*, vol. 42, pp. 297–346, 1990.

[78] M. Andreassen, B. F. M Woldbye, and S. K. Andersen, "Munin - a causal probabilistic network for interpretation of electromyographic findings," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 266–372, North Holland, 1987.

[79] K. Olesen, U. Kirulff, F. Jensen, B. Falck, S. Andreassen, and S. Andersen, "A munin network for the median nerve - a case study on loops," in *Applied Artificial Intelligence. Special Issue: Towards Causal AI Models in Practice* (H. E. Shrobe, ed.), pp. 346–379, San Mateo (CA): Morgan Kaufmann, 1989.

[80] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, L. H.P., and C. G.F., "Probabilistic diagnosis using a reformulation of the internist-1/qmr knowledge base," tech. rep., KSL-90-09, Knowledge Systems Laboratories, Stanford University, 1990.

[81] D.B.First, L. J. Soffer, and R. A. Miller, "Quick (quick index to caduceus knowledge); using the internist-1/caduceus knowldge base as an electionic textbook of medicine," *Computers and Biomedical Research*, vol. 18, pp. 137–165, 1985.

[82] D. McDermott, "A temporal logic for reasoning about processes and plans," *Cognitive Science*, vol. 6, pp. 101–155, 1982.

[83] T. Dean and D. McDermott, "Temporal data base managment," *Artificial Intelligence*, vol. 32, pp. 1–55, 1987.

[84] J. F. Allen, "Maintaining knowledge about temporal intervals," *Comminications of th ACM*, vol. 26, pp. 832–843, 1983.

[85] J. F. Allen, "Towards a general theory of actions and time," *Artificial Intelligence*, vol. 23, pp. 123–154, 1984.

[86] M. Kahn, *Model-based Interpretation of Time-oriented Data*. PhD thesis, University of California, San Francisco (CA), 1988.

[87] T. A. Russ, "Using hindsight in medical decision making," *Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care*, vol. 36, p. 589, 1989.

[88] C. Berzuini, "A probabilistic framework for temporal reasoning." Submitted to *Artificial Intelligence*, 1992.

[89] C. Langlotz, *A Decision-Theoretic Approach to Heuristic Planning*. PhD thesis, Stanford University, San Francisco (CA), 1989.

[90] I. Haimowitz, R. Patil, and p Szolovits, "Representing medical knowledge in a terminological language is difficult," in *Proceedings of the Annaual Symposium on Computer Application in Medical Care*, pp. 101–105, 1988.

[91] H. Levesque and R. Brachman, "Expressiveness and and tractability in knowledge representation and reasoning," *Computational Intelligence*, vol. 3, pp. 78–93, 1987.

[92] B. Hayes-Roth, "A blackboard architecture for control," *Artificial Intelligence*, vol. 26, pp. 251–321, 1985.

[93] P. Nii, "Blackboard systems," tech. rep., Knowledge Systems Laboratory, Stanford University, Stanford (CA), 1986.

[94] S. Quaglini, M. Stefanelli, G. Barosi, and A. Berzuini, "Anemia: An expert consulation system," *Computers and Bio-medical Research*, vol. 19, pp. 13–27, 1986.

[95] S. Quaglini, M. Stefanelli, G. Barosi, and A. Berzuini, "A performance evaluation of the expert system ANEMIA," *Computers and Bio-medical Research*, vol. 21, pp. 307–323, 1988.

[96] J. de Kleer, "An assumption-based tms," *Artificial Intelligence*, vol. 28, pp. 127–162, 1986.

[97] S. Quaglini, C. Berzuini, R. Bellazzi, M. Stefanelli, and G. Barosi, "Therapy planning by combining ai and decision theoretic techniques," in *Proceedings of the Second European Conference on Artificial Intelligence in Medicine*, (London), pp. 147–156, Springer-Verlag, 1989.

[98] B. Hayes-Roth, "Architectural foundations for real time performance in intelligent agents," in *Real-Time Systems* (J. C. Carbobell and T. M. Mitchell, eds.), pp. 593–624, Los Altos: Morgan Kaufmann, 1990.

[99] B. Hayes-Roth, R. Washington, R. Hewett, M. Hewett, and A. Seiver, "Intelligent real-time monitoring and control," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 243–249, 1989.