Deep and shallow models in medical expert systems

E.T. Keravnou and J. Washbrook

Abstract. In the context of medical expert systems a deep system is often used synonymously with a system that models some kind of causal process or function. We argue that although causality might be necessary for a deep system it is not sufficient on its own. A deep system must manifest the expectations of its user regarding its flexibility as a problem solver and its human-computer interaction (dialogue structure and explanation structure). These manifestations are essential for the acceptability of medical expert systems by their users. We illustrate our argument by evaluating a representative sample of medical expert systems. The systems are evaluated from the perspective of how explicitly they incorporate their particular models of expertise and how understandably they progress towards solutions. The dialogue and explanation structures of these systems are also evaluated. The results of our analysis show that there is no strong correlation between causality and acceptability. On the basis of this we propose that a deep system is one that properly explicates its underlying model of human expertise.

Key words: medical expert systems, deep model, explicitness of knowledge, human-computer interaction, dialogue structure, explanation structure, ABEL, CASNET, INTERNIST-I, MDX, MYCIN, NEOMYCIN.

1. Introduction

Medical expert systems have significantly contributed to the advancement of the technology of expert systems (Clancey and Shortliffe 1984). The diagnostic systems MYCIN, PIP, and INTERNIST-I are among the earliest developed expert systems. However, despite the fact that several of these medical systems have achieved high levels of performance, hardly any has progressed from the research laboratory into practical use.

Computer technology has been widely applied in medicine, in recent years for providing automated decision aids for clinicians (Young 1982). Medical expert systems are only a subset of the AI applications in medicine (Shortliffe et al. 1979). By and large, medical expert systems are diagnostic systems which, more often than not, do not make treatment recommendations. Some of the more recent medical expert systems, though, have the wider (and possibly harder to attain) objective of aiding in the complete management of patients. Diagnostic systems aim to pinpoint the cause of the abnormal findings in the patient. Patient management systems aim to recover the patient from an unhealthy state of affairs; this could involve making time-critical decisions (based on the status of the patient, all the possible causes of his problem and their respective implications), as well as the monitoring of the patient over a period of time. Isolating the cause of the abnormal findings is not the objective of a patient management system; its objective is the revocation or the prevention of undesirable effects of the likely causes. Treatment advisory systems lie between diagnosticians and patient managers. The objective of a treatment advisor is to narrow down on the cause of the problem as closely as a course of treatment would entail and to recommend the optimal treatment plan for the given patient. Treatment advisors may not necessarily follow up the progress of the patient like a fully fledged patient management system.

The majority of medical expert systems are consultative systems that need to engage in active interactions with humans. The weaknesses of the human-computer interaction in current medical expert systems is a prime factor in the medical community's scepticism towards such systems. It is fair to say that medical expert systems have not yet been accepted by their potential users.

There is a general agreement as to what is a *shallow* model of expertise, namely one that models expertise as a collection of if-then associations (rules). Most first generation expert systems are shallow. A recent advance in expert systems is the so-called *deep* expert system. Various definitions have been put forward for what is a deep model of expertise (Price and Lee 1988). The notions of causality, temporal reasoning, qualitative reasoning, reasoning from first principles, or reasoning from structure and function all figure in these definitions. The advocates of deep models claim that richer explanations, more adequate dialogue structures, and higher flexibility in problem solving yielding higher levels of performance accrue from this approach. Klein and Finin (1987) have proposed the following comparative definition of deepness: a model M is deeper than a model M' if M represents knowledge or is able to infer knowledge that is implicit in M'.

In this paper we develop a definition based on Klein and Finin and use it to rank some existing medical expert systems on a 'deepness' scale; we evaluate the human-computer aspects of these systems and compare our deepness ranking with the quality of interactions.

1.1 Analytical framework

Human-computer interaction encompasses a wide range of issues from the very technical to the highly abstract. From the perspective of this paper the relevant aspects of human-computer interaction are the *dialogue structure* and the *explanation structure*.

Dialogue structure covers the ordering of the questions raised by the system, the relevancy and comprehensibility of these questions, and the nature of the interaction in general (mixed-initiative for example). The number of questions raised is not in itself a sufficient metric for the quality of interaction. (Needless to say that asking too many questions, especially in time-critical situations, is unacceptable.) Human experts are able to home onto the problem with the minimum of information. The ability of experts to distinguish relevant from irrelevant information in a particular problem is an attribute of their expertise (Elstein et al. 1978).

Explanation structure covers the way each explanation is presented, the level and depth of explanation, and the model of the user upon which the explanation is based. Because the ultimate responsibilty for the treatment given to a patient rests with the medical practitioner, she/he should be able to understand the reasoning behind that treatment. Hence explanations are particularly vital in the context of medical expert systems. Yet the majority of medical expert systems do not provide an explanation facility (and some that do have been severely criticized). This must be a major reason for the lack of acceptance of these systems.

Figure 1 gives the top-level framework for an expert system. For medical expert systems, the case picture holds the data specific to the particular patient and the progressions towards the solution (i.e., generated hypotheses, conclusions, decisions, actions). In the context of diagnostic systems the case picture may be referred to as the diagnostic picture. The domain factual knowledge is general knowledge that should cover for any specific case in that domain. The reasoning knowledge is also by and large domain specific; it manipulates the domain factual knowledge and the case specific information to progress the solutions. The structuring and interactions of these three components, domain factual knowledge, reasoning knowledge and case picture, constitute the model of expertise incorporated in the given system. The nature of the human-computer interaction depends on this model (Keravnou and Johnson 1986).

In Section 2 we analyse the selected medical systems from two perspectives: (1) the 'explicitness' of the incorporated model and (2) the nature of the human-computer interaction. In the background of Figure 1, a model is explicit if: (a) the domain *reasoning knowledge* is understandable and such knowledge is not implicit in control constructs; (b) the domain *factual knowledge* is sufficiently differentiated into its types and such knowledge is not implicit in control constructs; and (c) the progress towards the solution is explainable at any intermediate point as well as at the end of the consultation.

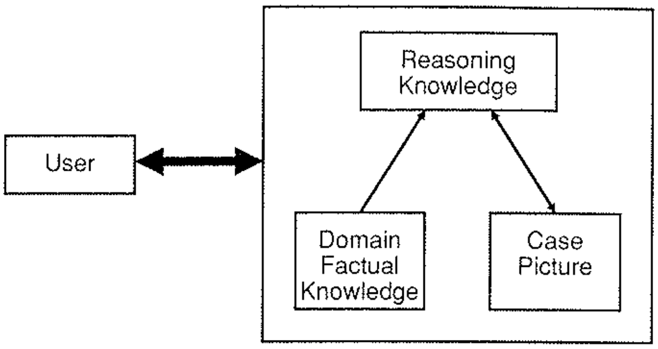


Figure 1. Expert system framework. A deep system must manifest the user expectations regarding a meaningful dialogue and appropriate explanations.

The analysis of the human-computer interaction is concerned with the two aspects: (i) dialogue structure and (ii) explanation structure. Some of the selected systems do not provide an explanation facility per se but rather a trace of their operations. In these systems the trace is taken to be the explanation structure.

2. Case analyses

We have selected six medical expert systems to discuss in the sequel: the diagnostic systems MDX, INTERNIST-I and NEOMYCIN; the treatment advisors CASNET and MYCIN, and the patient manager ABEL. To the best of our knowledge only the diagnostic component of ABEL has been demonstrated in a working system and thus, in this paper, we treat ABEL as a diagnostic system as well. The selected systems constitute a representative sample of medical expert systems.

2.1 MYCIN

2.1.1 Model of expertise

MYCIN (Shortliffe 1976) diagnoses certain antimicrobial infections and recommends drug treatment. The MYCIN framework is depicted in Figure 2.

Explicitness of domain factual knowledge

MYCIN's factual knowledge is by and large represented in

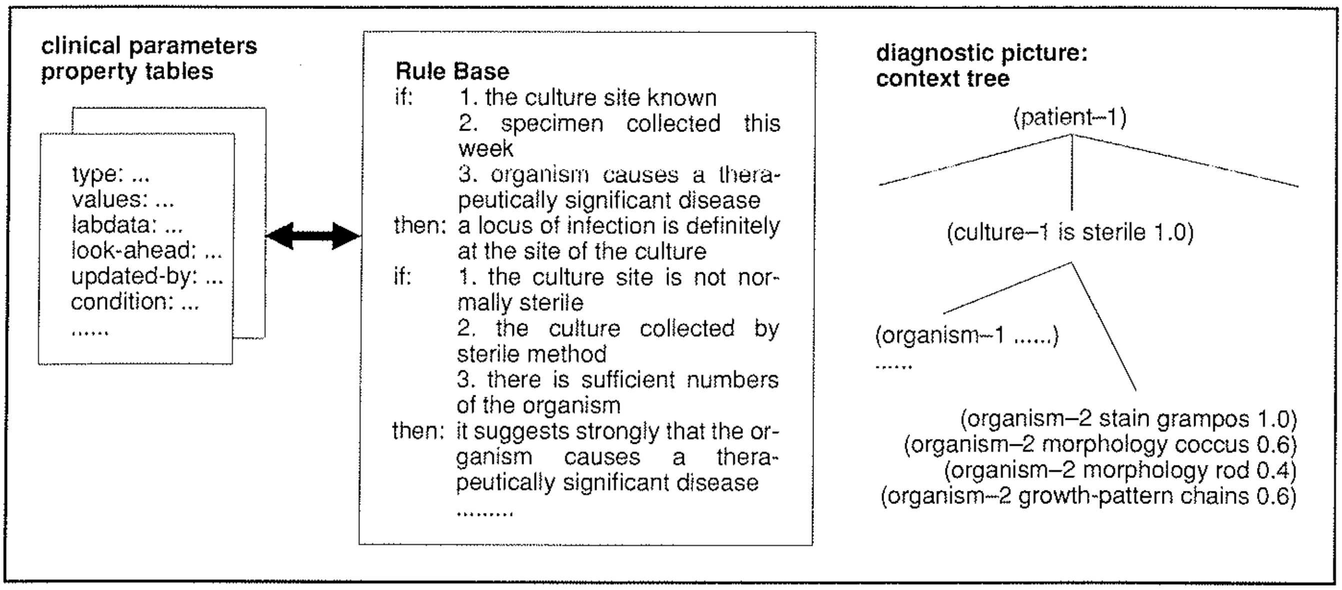

clinical parameters
property tables
type: ...
values: ...
labdata: ...
look-ahead: ...
updated-by: ...
condition: ...
......
Rule Base
if: 1. the culture site known
2. specimen collected this week
3. organism causes a therapeutically significant disease
then: a locus of infection is definitely at the site of the culture
if: 1. the culture site is not normally sterile
2. the culture collected by sterile method
3. there is sufficient numbers of the organism
then: it suggests strongly that the organism causes a therapeutically significant disease
....
diagnostic picture:
context tree
(patient–1)
(culture–1 is sterile 1.0)
(organism–1 ......)
......
(organism–2 stain grampos 1.0)
(organism–2 morphology coccus 0.6)
(organism–2 morphology rod 0.4)
(organism–2 growth-pattern chains 0.6)

Figure 2. MYCIN framework

terms of if-then associations or rules. Below we give a diagnostic rule and a treatment rule:

If 1) the infection is meningitis
2) the subtype of meningitis is bacterial
3) only circumstantial evidence is available
4) the patient is at least 17 years old and
5) the patient is an alcoholic

then there is suggestive evidence that diplococcus pneumoniae is an organism causing the meningitis

If the identity of the organism is pseudomonas

then I recommend therapy chosen from among the following drugs:
1. Colistin (0.98)
2. Polymyxin (0.96)
3. Gentamicin (0.96)
4. Carbenicillin (0.96)
5. Sulfisoxazole (0.64)

The diagnostic rules are highly stylized and interact implicitly through the sharing of antecedent or consequent clauses. Each treatment rule essentially associates an organism with the set of drugs which have been shown to be effective against the organism.

When a diagnostic rule is evaluated, its antecedent conditions **must** be evaluated in the given order. Referring to the example rule, establishing that meningitis is present in the patient must precede the establishing that its type is bacterial, something immediately obvious to (trained) humans. However, MYCIN's model does not allow for the explicit representation of etiological taxonomies. Looking further into the antecedent of this rule we see that the age of the patient must be established prior to asking whether the patient is alcoholic. World facts relevant to MYCIN's domain such as 'it is generally assumed that children are not alcoholics' are not explicit in the MYCIN model.

Once the diagnostic rule antecedent is stripped of its contextual (1–3) and screening (4) clauses, what is left is that alcoholism suggests that diplococcus penumoniae is an organism causing the meningitis. However, the logical basis (justification) for this association is not present in the model. Without this information it is not possible to see in which situations this association is violated (Dhar and Pople 1987). A MYCIN diagnostic rule is nothing more than an evolved pattern of reasoning that copes with the demands of ordinary problems leaving out 'unnecessary' steps.

In addition to the diagnostic rules, the model incorporates factual knowledge about clinical parameters such as the morphology and aerobicity of organisms in terms of property tables. Included properties specify the type and permitted values for the parameter. The two properties *updated-by* and *look-ahead* are not conceptual properties, but rather mechanical ones which simply provide indexes into the rule base. Another property, *condition*, specifies conditions which need to be established prior to asking the user questions about the value(s) of the parameter. Implicit in these questions are world facts such as 'sex male implies pregnancy absent', and common-sense principles such as 'if a class of things is absent then so is any subtype of it'.

Explicitness of reasoning knowledge

Perhaps the only domain reasoning knowledge given explicitly in the MYCIN model is that first you identify the offending organisms and then you prescribe the best treatment. This is given by MYCIN's *goal rule* which links the two aspects of MYCIN, namely diagnosis and treatment. For its diagnostic aspect MYCIN employs a domain independent (and thus very mechanistic) inference mechanism via two procedures, *monitor* and *findout*. These procedures do not embody any domain specific diagnostic strategies, and could be used in any rule-based system to chain backwards from some initial goal. The concept of a hypothesis is alien to MYCIN (Clancey 1986).

We said earlier that the rule antecedents must be evaluated in the order given. Implicit in the first two clauses of our example diagnostic rule is the reasoning strategy that we first establish a class of etiologies (meningitis) and then we proceed to refine this to a particular instance of the class (bacterial meningitis). The absence of an etiological taxonomy makes the explication of this strategy impossible. In this rule, clause number 3 establishes that only circumstantial evidence is available. Implicit in this is the reasoning strategy that evidence is either direct or circumstantial, and that different association strengths should be used in the presence of both direct and circumstantial evidence. In the presence of direct evidence a companion rule to the above allows the circumstantial evidence of alcoholism to be considered but gives it less weight. Thus in MYCIN the actual domain strategies are implicit in the control constructs employed, e.g., ordering of antecedent clauses and ordering of rule invocations.

The domain reasoning for treatment selection is represented in terms of Lisp procedures. These procedures embody heuristics such as 'use the minimum number of drugs' and 'no more than one drug from the same category should be used'. The treatment knowledge in MYCIN is therefore implicit in the given Lisp procedures.

Explicitness of case picture

In MYCIN the diagnostic picture consists of (*context*, *parameter*, *value*, *CF*) quadruples which are hierarchically organized through their contexts. The context tree registers all the patient-specific findings quite explicitly. This may seem paradoxical compared to the implicitness of the domain, factual and reasoning, knowledge in the MYCIN model. The understandability of the case picture is attributed to the systematic search regime applied through the monitor and findout and the absence of user volunteered information (see below).

2.1.2 Human-computer interaction

Dialogue structure

In MYCIN the user is not allowed to volunteer any information. The dialogue is entirely machine-initiated. MYCIN's search space is relatively small and thus exhaustive searching of it is possible. However, MYCIN's essentially blind search leads to unnecessary questions which, coupled with the need to re-enter all patient data anew for every consultation about the same patient, leads to a rather unacceptable situation. MYCIN tries to alleviate this by always asking more general questions than the current situation requires, e.g., instead of asking whether the morphology of organism-1 is rod, MYCIN would ask what is the morphology of organism-1. Similarly, subgoals are more general than the particular goal requires. Through this technique, MYCIN is trying to achieve a more organized and focussed approach to its diagnostic task. Lastly, with the condition property of parameters and the screening clauses in rules, MYCIN hopes to prevent irrelvant or incomprehensible questions. In spite of such (low level) control constructs, however, MYCIN's dialogue structure is not natural.

Explanation structure

The only explanations that MYCIN is capable of are in terms of the rule activations that took place in the particular consultation. Thus the quality of the explanations is dependent on the quality of the rules. As we have seen above, the justifications of rules as well as their structure (clause ordering) are not recorded. MYCIN's explanation structure is therefore not adequate.

2.2 CASNET

The *C*ausal *AS*sociational *NET*work (CASNET) embodies a model for the long-term management of diseases whose mechanism is well known (Weiss 1974). The CASNET model is illustrated in Figure 3.

2.2.1 Model of expertise

Explicitness of domain factual knowledge

In CASNET a disease process is modelled in terms of a causal network of dysfunctional states. The level of resolution used in the causal model is application dependent and must be the appropriate level from a diagnostic/prognostic perspective. It may well be that some of the states in a particular application cover a number of events.

Some of the dysfunctional states are *starting* states and some *final* states. Starting states are assigned prior frequencies denoting their likelihoods of occurring. The represented relationship between states is '*causes*', the relationship is multi-valued (or fuzzy) where the assigned value denotes the strength of causation. The severity of the disease increases as we move down a causal chain and the objective is to prevent the progression of the disease to a final state (provided it has not already progressed that far) by recommending the appropriate therapy regime. By modelling the *causes* relationship instead of its inverse *caused-by*, CASNET reasons forwards in time, thus enabling prognostic patient assessment.

Dysfunctional states are not observable entities but can be hypothesised on the basis of observations. Observations are *associated* with states; again this is a multi-valued relationship where the values specify the strength to which observations

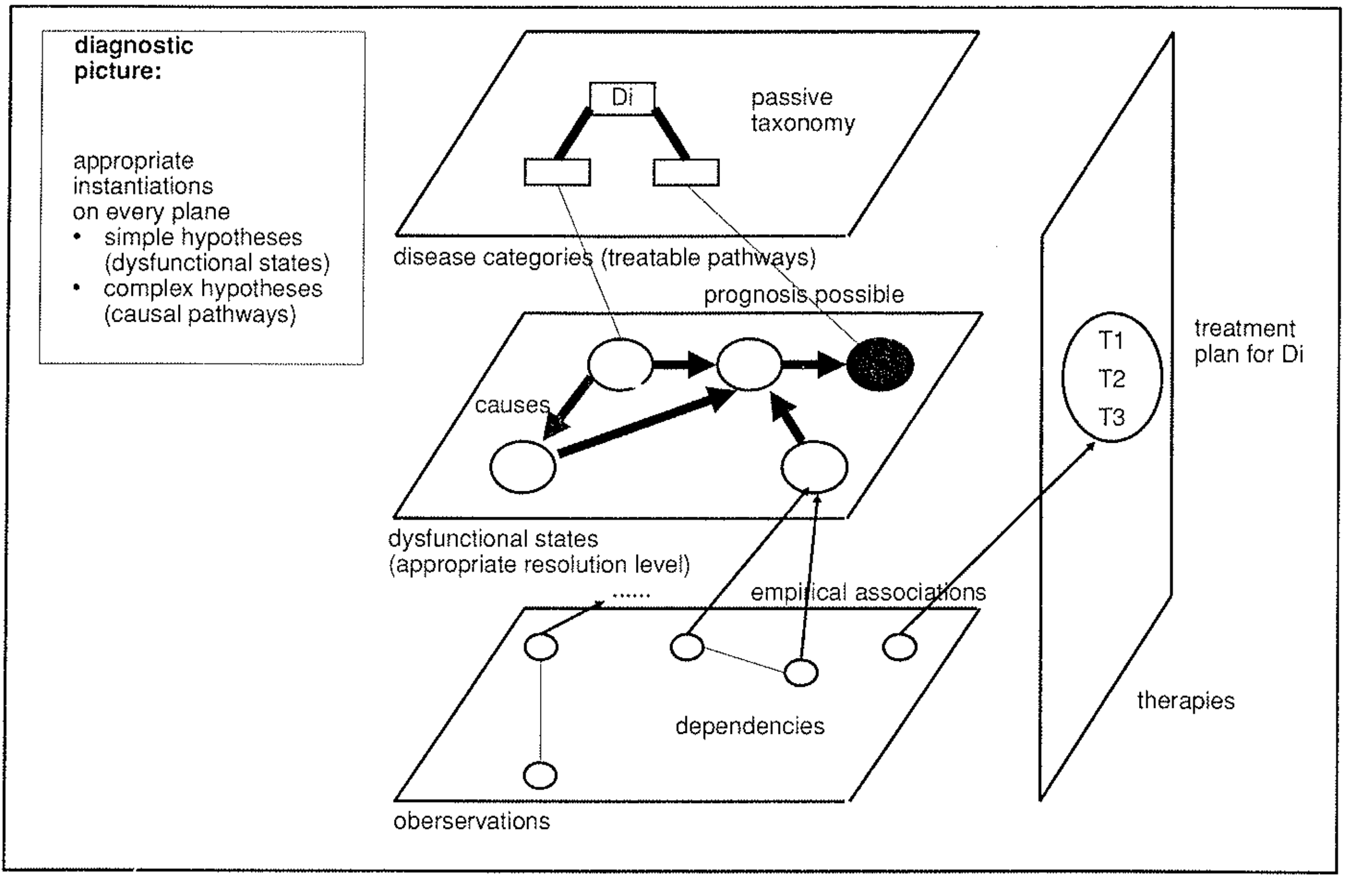

diagnostic picture:
appropriate instantiations on every plane
• simple hypotheses (dysfunctional states)
• complex hypotheses (causal pathways)
Di
passive taxonomy
disease categories (treatable pathways)
prognosis possible
causes
dysfunctional states (appropriate resolution level)
empirical associations
dependencies
oberservations
T1
T2
T3
treatment plan for Di
therapies

Figure 3. CASNET framework

provide positive or negative evidence for states. Observations are conceptually different from states, and thus in the CASNET model they occupy a plane of their own. Dependencies between observations are represented within their plane and these dependencies enable the derivation of information from known information and also provide question orderings.

Some of the dysfunctional states are designated as *treatable* states. Disease categories correspond to final states and thus include all the causal pathways (from starting states) that lead to their particular final states. Intermediate treatable states on these pathways define disease subcategories. This knowledge is represented on yet another plane. This plane is related to the plane of therapies. (The observations, states and disease categories planes collectively represent the diagnostic knowledge.)

Disease categories are related to therapy plans which cover for all their subsumed subcategories. Therapy plans are also associated with those observations that provide indications or contraindications for treatments included in the plan.

The CASNET model indeed provides a fine differentiation of domain concepts and their interrelationships.

Explicitness of reasoning knowledge

The CASNET diagnostic process aims to identify the causal pathways which are currently operative in the patient. This is done by repeatedly collecting oberservations, hypothesising dysfunctional states and then hypothesising the most likely causal pathways. A state is dynamically assigned two measures: a *confidence factor* (belief measure) derived from the instantiated observations related to it and a *weight* (likelihood measure) derived from contextual evidence, i.e., other instantiated states that are causally related to it. Once no further questioning is considered useful diagnostic subcategories are identified on the most likely causal pathways and treatments are recommended. The disease categories plane does not play an active role during the diagnostic process.

However, as the reasoning knowledge is entirely coded in terms of Fortran subroutines, this knowledge is not sufficiently explicated.

Explicitness of case picture

The case picture consists of the portions of the planes which have been instantiated for the particular case. What is interesting here is the presence of two hypothesis spaces: (1) the space of individual dysfunctional states constituting simple hypotheses, and (2) the space of causal pathways constituting complex hypotheses, which may be an infinite space and which is dynamically generated.

The causality relationship is central to the complex hypothesis space. The components of complex hypotheses are

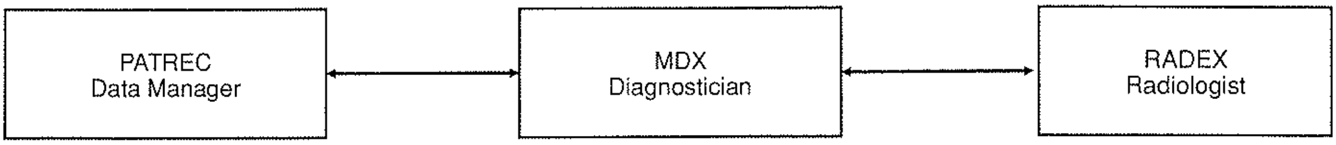


Figure 4. (a) MDX and its satellite systems. Each system consists of a taxonomy of cooperating experts.

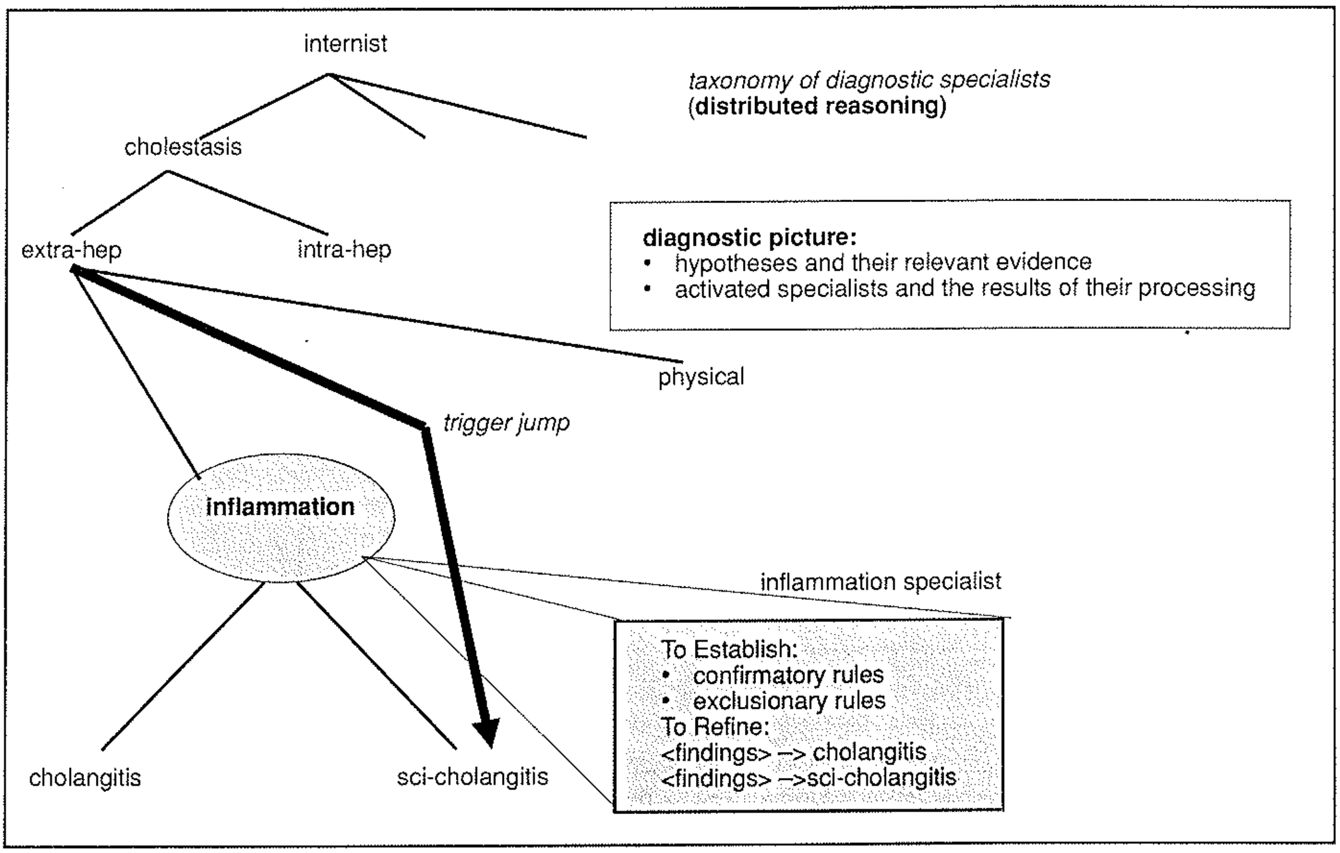


Figure 4. (b) MDX framework

constructed in parallel rather than sequentially (cf. INTERNIST-I). This provides for diagnostically complete and coherent hypotheses.

The diagnostic picture therefore holds all the partial solutions suggested by the observations and thus shows how well a partial solution explains the observations. This explicit representation of diagnostic possibilities gives a complete and meaningful progression towards the solution.

2.2.2 Human-computer interaction

Dialogue structure

The dialogue permitted by the CASNET model is mixed-initiative; the user can volunteer information and the system can ask questions to elicit additional information.

The domains where the CASNET model has been applied are rather narrow and in a sense comparable to MYCIN's domain. However, unlike MYCIN which performs an exhaustive search of its space, CASNET abduces hypotheses and then selects and applies an information acquisition strategy which gives it a much more natural dialogue structure. Questions aim at confirming the presence of states which are suggested either by observations or other established/suggested states. The answers to questions are used, in conjunction with a fixed formula, to determine *confidence factors* for their associated states. These confidence factors are then translated in a fuzzy way into *status* values (unknown, denied, present). The *causal strengths* associated with arcs between states are combined with the status values to give each state a *weight*, denied states (for example) being given low weights. These weights, derived from the causality relationship, are used to direct information acquisition in a dynamic, focussed way. The causality relationship also determines the most likely complex hypothesis.

Explanation structure

Although the domain factual knowledge is quite explicit, the reasoning knowledge is embedded in Fortran routines. One would, therefore, rightly predict that there would be limits to the kinds of explanations that CASNET could be capable of.

For example, it may not be easy to get run time explanations, i.e., what the system is currently doing and why in terms understandable to the system users.

In fact CASNET can only provide explanations regarding its conclusion at the end of a consultation: (1) it can explain why a given state was confirmed or denied by displaying the evidence relevant to it, i.e., the relevant associational links which were instantiated; (2) it can explain why a given causal pathway was selected; and (3) it can explain how conflicts of evidence were resolved.

Although the above explanations are very useful to have, certainly they are not sufficient on their own. The inability to dynamically provide question justifications is indeed a serious omission.

2.3 MDX

MDX performs diagnoses in the liver syndrome cholestasis (Chandrasekaran and Mittal 1983). Critical to the overall performance of MDX are its two auxiliary systems, PATREC and RADEX (see Figure 4 a). PATREC manages the patient data and performs 'intelligent' data retrieval and inference. RADEX is a radiological consultant which interprets various kinds of imaging data. Each of these three systems is structured in the same way. This framework is illustrated in Figure 4 b through the MDX system. The designers' objective was to explicate the *conceptual structure* of the domain knowledge (Chandrasekaran and Mittal 1982); this conceptual structure reflects the uses of the knowledge in performing the particular global task effectively.

2.3.1 Model of expertise

Explicitness of domain factual knowledge

The diagnostic knowledge is *distributed* among the nodes of a diagnostic taxonomy (see Figure 4 b). Each node represents some *specialist*, e.g., an internist. The MDX diagnostic taxonomy plays a very active role in relation to the passive nature of CASNET's disease taxonomy.

Each diagnostic specialist has knowledge about establishing the presence of a problem in its area of specialisation (confirmation and exclusion criteria), and in the case of a non-primitive specialist, knowledge about refining that problem (the latter knowledge can be activated if the relevant diagnostic possibility has not been ruled out); this knowledge may include refinement suggestions for any subsequent level in the diagnostic taxonomy and not just for the immediate successors (multi-level refinement jumps in the context of top level specialists are *triggers*). The knowledge comprising a diagnostic specialist is either represented declaratively, e.g., in terms of rules, or in terms of procedures; in either case this knowledge is activated through some interpreter.

As mentioned above, PATREC exhibits the same conceptual organisation as MDX. The nodes of PATREC's taxonomy represent specialists on medical concepts, e.g., a specialist on the concept of drugs. PATREC has deep knowledge of medical concepts, such as the units of measurements used and the relevant laboratory tests for eliciting findings on these concepts. In addition, PATREC has an understanding of rudimentary temporal aspects. This is demonstrated through the following example taken from Chandrasekaran and Mittal (1983).

Suppose that the patient data include the following: (a) halothane was administered at cholecystectomy, (b) bilirubin 12.2 three days after surgery, and (c) patient had pruritus a week later. And suppose that the cholestasis diagnostic specialist wants to evaluate the following (refinement) rule: If jaundice onset within a week after surgery and pruritus developed after jaundice, then consider post-operative cholestasis caused by anaesthetics.

The MDX specialists may invoke PATREC when evaluating rule premises. In the above example PATREC's reasoning is outlined as:

Data abstractions

1. Cholecystectomy is a type of surgery (generalisation)
2. Bilirubin 12.2 means bilirubin elevated (qualitative abstraction)
3. Bilirubin elevated implies jaundice present (data dependency)

Temporal reasoning

4. 3 days is less than a week and jaundice appeared the same time as bilirubin
5. Pruritus developed after jaundice (a week after the time at which bilirubin was elevated).

From the above example we see that PATREC possesses a rich knowledge on medical concepts that enables it to generalise or restrict a finding (through the conceptual taxonomy), derive dependencies on it, and translate quantitative values into qualitative ones (definitional links). The explication of definitional, implicational and generalisation links is necessary for generating suitable abstractions over the patient data.

Explicitness of reasoning knowledge

MDX is a diagnostic system; it does not make any treatment recommendations. The objective of MDX is to identify the primitive diagnostic concepts (terminal nodes in the diagnostic taxonomy) which are present in the patient. The diagnostic strategy is one of Establish and Refine, although this strategy has not been abstracted in the system as such (see below).

During a consultation control is initially passed to the topmost specialist, *internist* say. Other specialists are activated by a predecessor specialist through message passing. Messages include 'establish yourself', 'establish yourself and then refine', etc. A specialist operates on the message from its predecessor and returns the results of its operations to the calling specialist. If, for example, it fails to establish itself, it indicates so. This message passing is governed by rules that correspond to practices adopted by the human specialists (Chandrasekaran et al. 1979). For instance, no lateral calls are allowed; the calls are usually from a specialist to a subspecialist and vice versa.

In order to execute a request from another specialist, the subspecialist concerned activates the appropriate procedures attached to it. This is why the reasoning is said to be *distributed*; there is no centralised control and no explicit statement of the strategies used in general (abstract) terms. The trace of messages between the various specialists explicates the reasoning chain that took place in a particular consultation. How understandable this trace of messages is would depend on how obvious the justifications for the various exchanged messages are. These justifications are embedded in the procedures attached to the specialists. Thus for a deeper understanding we need to appreciate the logic of these procedures. The explicitness of MDX's reasoning knowledge, therefore, by and large depends on how explicit the specialist procedures are. The general domain strategies and their justifications are not given explicitly.

Explicitness of case picture

The case picture lists all the activated hypotheses/conclusions (diagnostic concepts) with their associated evidence, both positive and negative. Although not evident from the literature, we would guess that the trace of the specialist activations and the results of their operations are also registered in the case picture. The patient record (including historical information), on secondary storage, is managed by PATREC. Since all information on the patient is available at any time during a consultation, for all purposes the patient record is a component of the case picture.

2.3.2 Human-computer interaction

Dialogue structure

The user can volunteer information initially and the subsequent order of information requests depends on the order of specialist activations. A question either aims to establish some diagnostic concept or to refine it. In this respect the refinement suggestions are very critical. If they form accurate focusing heuristics, backtracking would be minimal and thus the questions raised would be relevant. The presence of the diagnostic taxonomy ensures (in a natural way) that general concepts are established prior to their specialisations and that the pursued hypotheses should be as specific as the data imply (since a specialist can invoke a subspecialist many levels down in the diagnostic taxonomy).

As seen above, the PATREC system plays a crucial role in managing the patient information and reasoning intelligently with it; thus this auxiliary system contributes much towards achieving a natural dialogue structure.

Explanation structure

MDX dynamically generates and displays a trace of its activities, i.e., the sequence of activated specialists and the operations they need to perform. Any conclusion reached is displayed with the evidence that led to it. Such a trace is very useful indeed and probably sufficient for some users. This explanation is entirely machine-initiated (the user cannot ask for any other explanations) and it is at a rather gross level. For example, it may not be obvious from the displayed message why a particular specialist is invoked (see above). This justification is implicit in a procedure attached to the invoking specialist. In theory, MDX could provide 'deeper' explanations by displaying these procedures; the quality of these explanations would therefore depend on the understandability of these procedures. The justifications of information acquisition requests are in terms of such procedures as well.

2.4 INTERNIST-I

INTERNIST-I is a diagnostic system for internal medicine (Miller et al. 1982). It is the largest AI in medicine program; its knowledge base covers 80% of internal medicine. The INTERNIST-I framework is depicted in Figure 5.

2.4.1 Model of expertise

Explicitness of domain factual knowledge

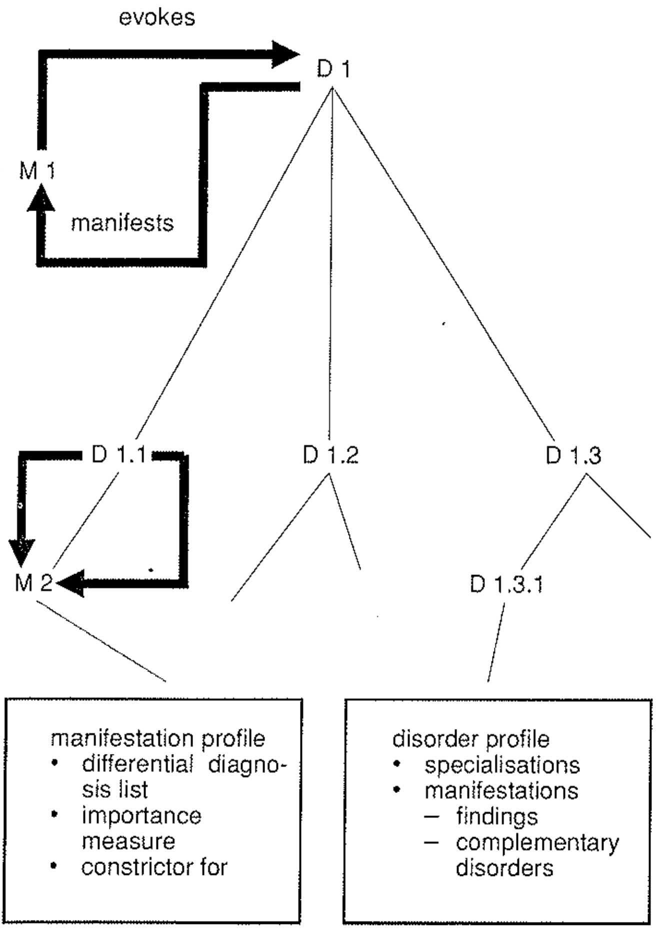
In INTERNIST-I there are essentially two domain concepts: disorders and manifestations. Disorders constitute the hypothesised entities while the concept of a manifestation is used ambiguously to denote findings as well as diseases related to some disease process (e.g., through causality). Manifestations are related to disorders via *evokes* associations. Evokes associations are multi-valued where the specified values denote the strengths to which the manifestations suggest the particular disorders. The inverse relationship to evokes is also explicitly represented; disorders are associated with their manifestations through a *manifest* relationship. Again, manifest associations are multi-valued where the values denote how strongly the presence of the disorder implies the presence of the associated manifestation. Thus evocative associations represent sufficiency measures and manifest associations embody necessity measures (how frequently a disorder exhibits the given manifestation). Disorders are related into a taxonomic structure where the manifestations of a disorder class are those shared by its specialisations. The INTERNIST-I taxonomy plays a role in the diagnostic process (but not as active as the MDX taxonomy of diagnostic specialists). The disorder taxonomy enables the explication of a special kind of relationship, the *constrictor* relationship; this is a strong association between a set of (usually easily observable) findings and a class of disorders, e.g., jaundice strongly suggests a liver problem; hence the observation of jaundice would constrict the investigation on liver diseases.

Finally, manifestations are associated with *importance* measures that indicate how important it is that the final diagnosis accounts for their presence in a patient.

Explicitness of reasoning knowledge

The objective behind INTERNIST-I was to build a system that modelled the reasoning of internists. The first version of INTERNIST-I was called DIALOG, an acronym for diagnostic logic.

In INTERNIST-I the reasoning knowledge is separated and abstracted from the factual knowledge in terms of Lisp procedures (cf. CASNET). There are procedures that generate hypotheses, and procedures embodying various information acquisition strategies (rule-out, pursue, etc.). Thus, unlike MDX where the reasoning knowledge is made specific and distributed among the specialists concerned, the INTERNIST-I reasoning is centralised. However, this reasoning is essentially embedded in the Lisp procedures that implement it.



diagnostic picture:

- current problem to be resolved
- concludes hypotheses (& manifestations accounted by them)
- unexplained important manifestations

Figure 5. INTERNIST-I framework

Explicitness of case picture

A case picture holds: the concluded hypotheses, each being associated with the manifestation accounted by it; the active hypotheses; and the unexplained manifestations. The picture is completed when every important manifestation is covered by some concluded hypothesis.

Although some of the manifest associations in fact represent causality associations between disorders, such causality relationships play a very secondary role in a diagnostic process in comparison with the CASNET model. INTERNIST-I does accept the presence of multi-disease illness but unlike CASNET it does not generate complex hypotheses. Instead it deals with simple disorder hypotheses, promoting the consideration of active hypotheses which are somehow causally related to concluded hypotheses. The complete complex explanation is therefore only apparent at the end of the consultation and not at intermediate stages (sequential rather than parallel reasoning).

2.4.2 Human-computer interaction

Dialogue structure

The dialogue is mixed-initiative; the user volunteers information initially as well as at subsequent stages when the system asks questions.

The system asks questions when it is pursuing a hypothesis, when it is trying to rule out some hypotheses, or when it attempts to differentiate between sets of hypotheses. These strategies emulate heuristics employed by human diagnosticians. The set of hypotheses to be resolved (current problem) are determined through a scoring heuristics and a partitioning heuristics. As soon as a new item of information is obtained, active hypotheses are rescored and repartitioned. This may lead to rapid changes in focus with a corresponding question sequence that diverges from that of human diagnosticians. This weakness of the INTERNIST-I system is not necessarily a consequence of the reasoning knowledge being implicit but rather a consequence of an inaccurate integration of the relevant heuristics. However, had the reasoning knowledge been more explicit and thus more understandable, the potential for its extending would have been enhanced.

The disorder taxonomy and the constrictor associations play an important role in reducing the number of unnecessary questions by focusing the diagnostic process at the initial stages; exhaustive searching in a space of the size of INTERNIST-I's is certainly not possible.

It should also be noted that INTERNIST-I's sequential reasoning in the case of multi-disease problems is bound to generate a dialogue structure that does not resemble that of physicians. The problem here is essentially the inappropriate explication and reasoning about the causality relationship (this drawback was addressed by the successor system CADUCEUS which combines a taxonomy of diseases with a causal network of pathological/disease states; by and large, though, CADUCEUS remains a paper system).

Explanation structure

INTERNIST-I displays a trace of its activities but no user-initiated explanations are possible. Through this trace the user knows which are the current conclusions, which manifestations have so far been explained, which are the active hypotheses and what strategy is used to acquire additional information (for resolving the current problem). The reasons for selecting the particular strategies are not disclosed and neither are the beliefs in the hypotheses explained. The trace is useful but insufficient since much of the system reasoning is hidden. Given the procedural representation of the system reasoning knowledge it would not be possible to provide more adequate explanations.

2.5 ABEL

The designers of ABEL aimed to build a manager for patients suffering from electrolyte and acid-base disturbances. From the available literature it appears that only the diagnostic aspect of ABEL has been implemented (Patil 1981) and hence this is what we discuss below. This model is given in Figure 6.

2.5.1 Model of expertise

Explicitness of domain factual knowledge

In ABEL the notion of causality is exploited in several ways: to organise the patient facts and disease hypotheses, to deal with the effects of more than one disease present in a patient, and to provide the basis for explanations. The causal relationship is given a rather novel interpretation, as a multivariate relation between various aspects of the cause and the effect, taking into account the context and the assumptions under which the causal link is being instantiated (see Figure 6 a). This interpretation is richer than CASNET's interpretation as the likelihood of observing the effect given the realization of the cause. In ABEL causal links are considered to be objects in their own rights giving the system the capability of hypothesising the presence or absence of a causal link between two realized nodes. Another important advantage of this is that the separate effects of multiple causes can be dynamically combined into one effect.

ABEL's knowledge-base consists of a causal network of nodes representing the domain of acid-base and electrolyte disturbances at the pathophysiological level. Through the mechanism of *focus links* and *focus nodes*, the system can dynamically generate abstractions of portions of this network at the clinical (phenomenological) level via an intermediate level. The 'causal' links at the clinical level span a number of causal pathways at the pathophysiological level.

The ABEL model, therefore, combines detailed pathophysiological knowledge with abstract clinical knowledge. It seems that competent physicians are able to reason at multiple levels of abstraction and to shift from one level of description to a more detailed or less abstract description level. The pathophysiological knowledge is for a more accurate attribution of findings (thus for the proper understanding of a difficult case) and the clinical knowledge is for focusing, by yielding a better exploration (global view) of the search space. The ability to dynamically generate abstractions of the search space is unique in ABEL amongst medical expert systems. Probably it should be noted that ABEL's application domain is considerably narrower than the majority of medical expert systems. (CADUCEUS, INTERNIST-I's successor, attempted to do something similar for internal medicine; as already pointed out, CADUCEUS is a paper system, however.)

Explicitness of reasoning knowledge

From a static perspective, the ABEL reasoning knowledge is embedded in Lisp procedures. Dynamically, however, ABEL explicates the reasoning involved in information acquisition in terms of a tree structure, the information acquisition plan. ABEL uses a set of information acquisition strategies that discriminate between a set of alternatives (similar to INTERNIST-I's strategies). The information acquisition plan is generated by repeatedly applying instances of the information acquisition strategies (pursue, explore, rule-out, discriminate, group-and-differentiate) to simpler and simpler subproblems of the goal problem which is to differentiate the causal hypotheses (see below). The non-terminal nodes of this tree specify instantiations of information acquisition strategies and the terminal nodes specify questions. Thus, the information acquisition plan explicates the rationale behind each question.

Explicitness of case picture

Like CASNET, ABEL has simple hypotheses (in terms of nodes and causal links) and complex hypotheses (in terms of causal chains that completely and coherently explain the patient problem). The case picture holds the various causal hypotheses with their promise scores and the current plan for information acquisition. However, it does not include the justification for selecting a particular acquisition action at a given point.

2.5.2 Human-computer interaction

Dialogue structure

The dialogue is mixed-initiative; the user volunteers information and the system asks questions. The information acquisi-

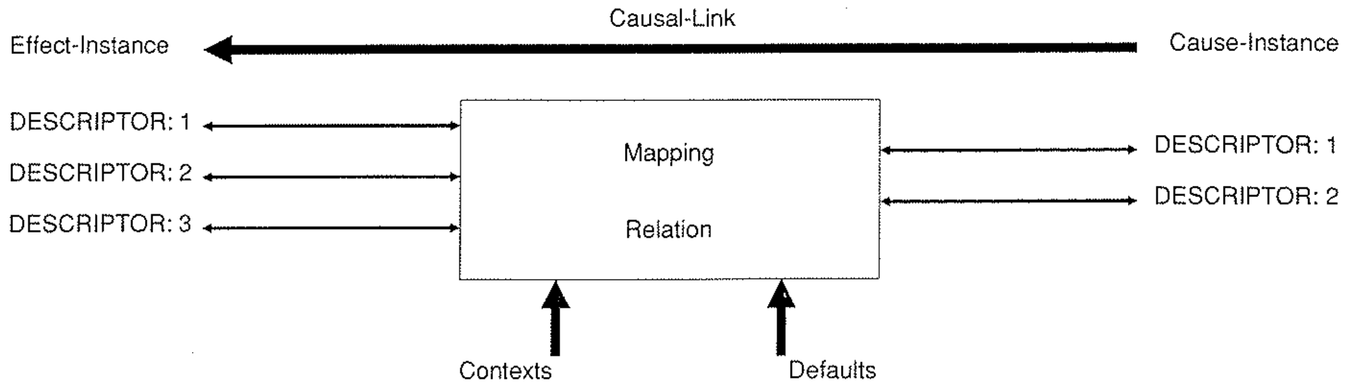
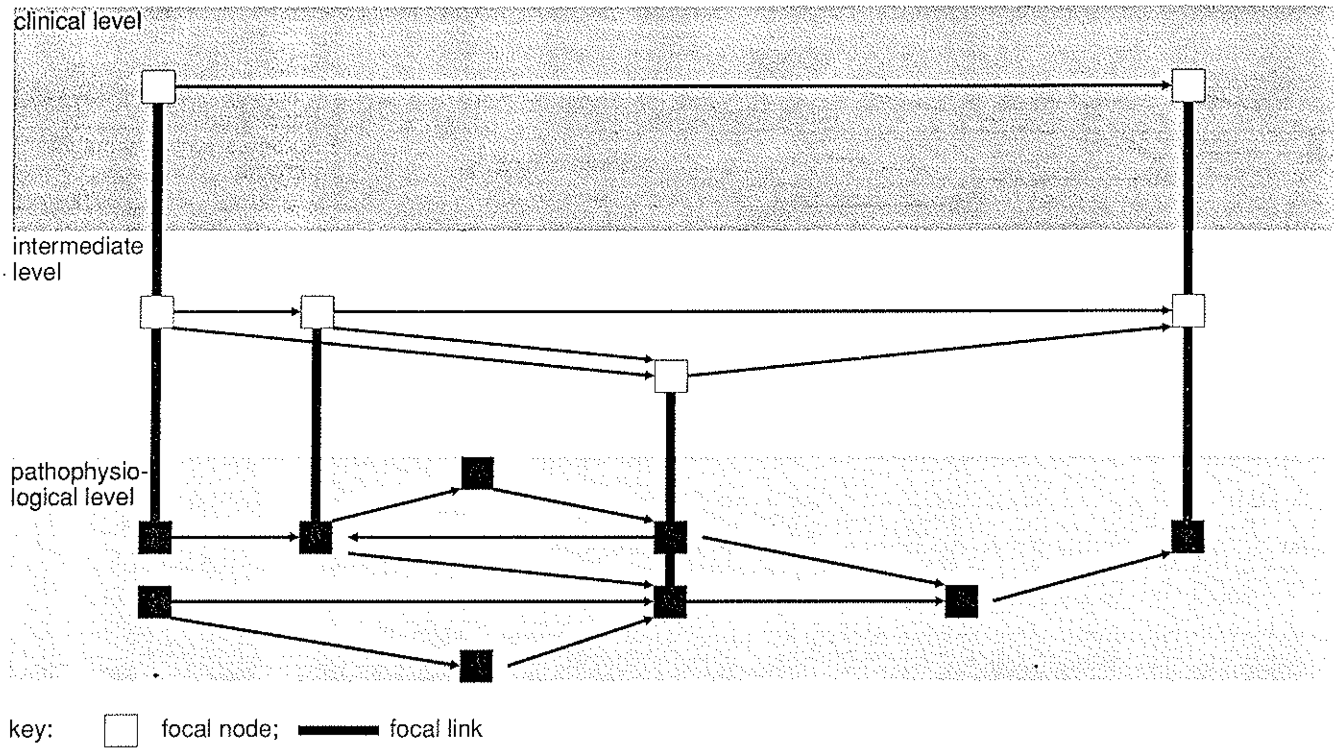


Figure 6. (a) Schematic description of a causal link.



diagnostic picture:

- complex hypotheses (patient specific models with their diagnostic closures)
- plan for information acquisition

Figure 6. (b) ABEL framework

tion plan plays a critical role in the quality of the incurred dialogue. Unlike INTERNIST-I which asks a single question and then reevaluates its hypothesis space, ABEL has a plan for information acquisition that aims to achieve a clinically meaningful and focussed pursuit of diagnosis. The causality relation and its interpretation are essential in the construction of the plan. Thus, 'causal' knowledge (at multiple levels of abstraction) can yield a more meaningful dialogue structure.

Explanation Structure

ABEL registers its reasoning regarding information acquisition in a tree, where a non-terminal node represents a diagnostic problem (i.e., a set of possibilities to be resolved) together with the differentiation strategy instantiated for this problem. Branches from the node lead to the means for achieving the corresponding strategy instantiation. Terminal

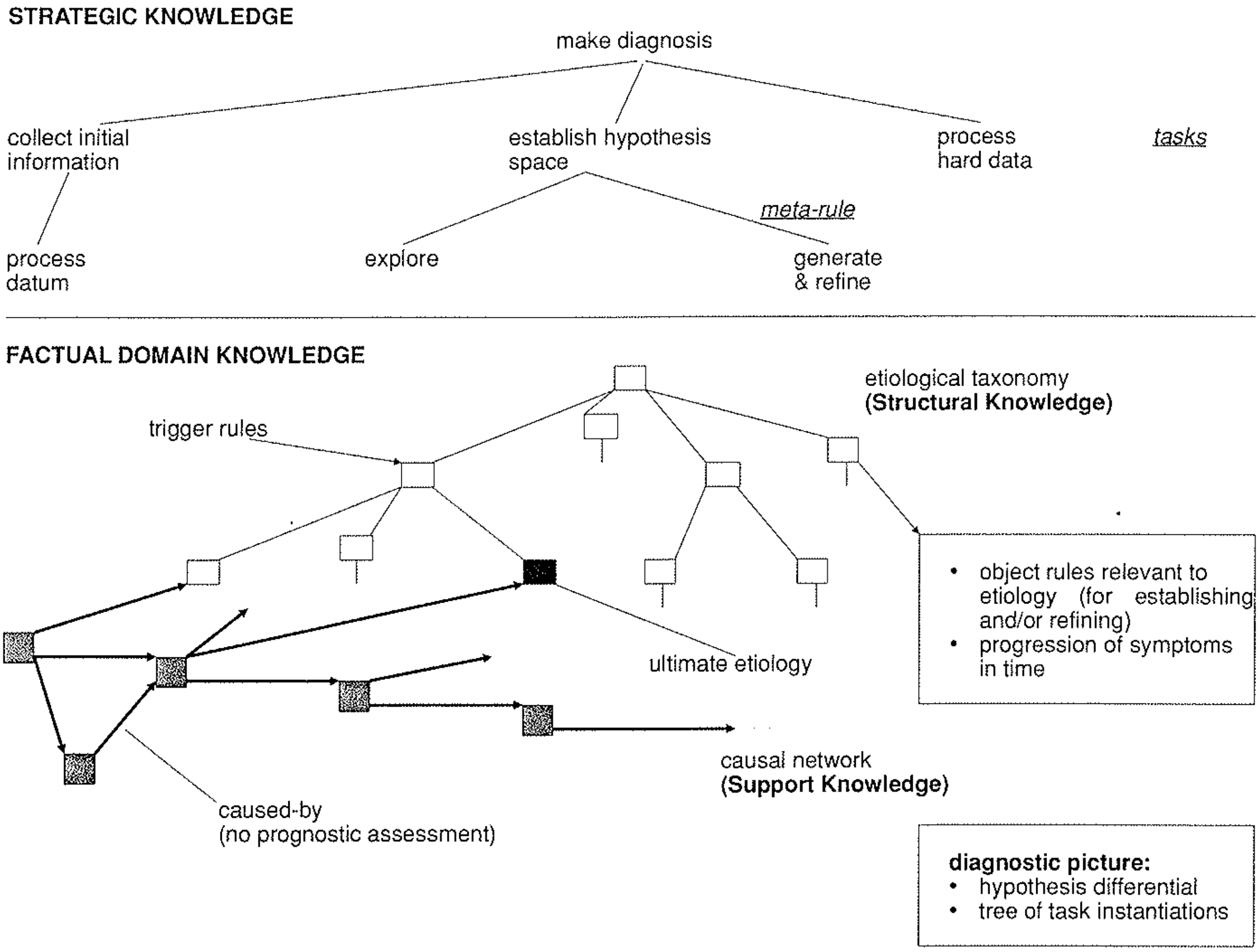


Figure 7. NEOMYCIN framework. Data dependencies represented in terms of screening rules

nodes represent questions. Thus the rationale behind a question can be obtained by traversing the tree upwards from the node representing the question. Although this tree registers useful information such as ABEL's expectations about the information being sought and how this information relates to the hypotheses under consideration, it does not explicate the reasons for selecting one information acquisition strategy rather than another; of course, it may be argued that such reasons are intuitively obvious, although there may be subtleties which are not intuitively obvious.

In addition, ABEL can provide English translations of the causal hypotheses at any given level of abstraction. Throughout a consultation it provides a trace of its activities (current hypotheses and their scores) and dynamically communicates the construction of the information acquisition tree. Although the trace and the explanations cover for much of ABEL's reasoning, albeit at a gross level, some of the reasoning is completely hidden from the user. For example, the user does not know what heuristics ABEL is using when scoring hypotheses or how the complex hypotheses are derived. The user is only told of the reasoning involved in differentiating between the competing hypotheses. Not being able to communicate to the user the rationale behind critical decisions is a serious omission.

2.6 NEOMYCIN

NEOMYCIN is a MYCIN derivative. The designers' objective was for a system that provided an efficient basis for teaching diagnostic reasoning and interpreting student behaviour (Clancey and Letsinger 1981). Figure 7 gives the NEOMYCIN framework.

2.6.1 Model of expertise

NEOMYCIN attempts to explicate the types of knowledge which are implicit in MYCIN. These types are referred to by Clancey as structural, strategic and support knowledge.

Table I. Summary of system models

	Domain factual knowledge types	Domain reasoning knowledge	Case picture
MYCIN	inferential associations between assertions	implicit	assertions organized in a context tree
CASNET	causal network (central structure allowing prognostic assessment) observations disease categories treatment plans	embedded in Fortran routines (centralised)	simple hypotheses complex causal hypotheses
MDX	diagnostic taxonomy knowledge about establishing and refining diagnostic possibilities general knowledge on medical concepts	distributed among the diagnostic concepts (specialists) either declaratively (inferential associations) or procedurally abstract strategies not explicit	activated diagnostic concepts messages exchanged between the diagnostic specialists
INTERNIST-I	manifestations disorders disorder classes complementary relations between disorders	embedded in Lips procedures (centralised)	concluded hypotheses unexplained manifestations current problem area
ABEL	rich interpretation of causal link multilevel abstractions of causal knowledge	embedded (statically) in Lisp procedures, but ... *some* justifications for information acquisition questions are dynamically explicated in a tree-structured information-acquisition plan abstract strategies implicit	complex, coherent hypotheses current plan for information acquisition
NEOMYCIN	etiological taxonomy triggers observations inferential associations disease process knowledge causal network (reasons backwards in time)	reasoning strategies abstracted and made explicit (declarative representation)	hypothesis differential diagnostic plan (task instantiations)

Explicitness of domain factual knowledge

NEOMYCIN's domain of expertise is primarily that of MYCIN, although NEOMYCIN's organisation of knowledge facilitates its extension to cover for other disorders as well.

NEOMYCIN includes an etiological taxonomy (structural knowledge). Etiologies provide the contexts for object rules: MYCIN rules stripped of their contextual and screening clauses (see Section 2.1.1). Object rules associate findings to etiologies. These associations are not of a causal nature; the causal chains justifying these associations are stored as canned text for explanation purposes. Etiologies are also represented as disease processes in terms of symptom progressions in time. In addition, *trigger rules* associate findings to those etiologies, which the findings suggest strongly. Abstractions and restrictions on findings and relevant world facts in general are represented in terms of *screening rules.*

The NEOMYCIN model also includes a causal network linking observations to etiologies via dysfunctional states. Unlike CASNET and ABEL, the represented relationship is *caused-by.*

The causal network is essentially for reasoning backwards in time rather than for prognostic assessment. Thus, the NEOMYCIN model provides two routes to an etiology: via the etiological taxonomy or via the causal network. The latter represents more detailed knowledge than the former. This situation is similar to ABEL's clinical and pathophysiological level, but the triggering mechanism is absent in ABEL. The etiological taxonomy is for focusing and for a more global exploration of the search space (trigger associations usually involve classes of etiologies), and the causal network is for a more accurate attribution of findings and the handling of difficult cases. This arrangement obviously provides for higher flexibility in problem solving and by implication a higher level of performance.

Explicitness of reasoning knowledge

The representation of the reasoning knowledge is really what makes NEOMYCIN stand apart from the other medical systems included in this study.

The reasoning knowledge is abstracted from the domain factual knowledge and is declaratively represented in terms of tasks and meta-rules (the NEOMYCIN meta-rules are semantically different from the MYCIN meta-rules). A task is associated with a set of meta-rules which model the means for achieving that task. Meta-rule consequents represent subtasks, and their antecedents define conditions which must be true for selecting the given subtasks. Thus, tasks are linked via the meta-rules into a hierarchical structure; the root task is 'make-diagnosis'. Tasks and meta-rules are given in parameterised form. Non-terminal tasks are essentially control tasks and terminal tasks perform factual knowledge manipulations and ask questions. Non-terminal tasks have termination conditions.

Explicitness of case picture

The case picture holds the *differential* (activated hypotheses) and the case findings. In addition, it contains the dynamically generated diagnostic plan. This is a tree structure whose nodes give the task instantiations for the particular case. A non-terminal task instantiation is achieved by repeatedly selecting and applying meta-rules associated with the task until its termination condition is satisfied. A meta-rule application results in the instantiation of a subtask. Thus, the diagnostic plan registers completely the reasoning for the particular consultation.

2.6.2 Human-computer interaction

Dialogue structure

NEOMYCIN has alleviated significant problems with MYCIN's dialogue structure. The user may volunteer information initially as well as at subsequent stages, and this information is used to constrain the hypothesis space. The etiological taxonomy and the trigger associations play an important role in this respect. Hence the NEOMYCIN taxonomy is an active one. The screening rules also play an important role in ensuring a more intelligent dialogue structure.

Table II. Summary of human-computer interaction

	Dialogue structure	Explanations structure
MYCIN	entirely controlled by system	replay of rule activations
CASNET	"mixed initiative" focus directed by the "weights" (values for the dysfunctional states)	can only explain its conclusions
MDX	"mixed initiative" focus determined by diagnostic specialists	trace of actions (system initiated)
INTERNIST-I	"mixed initiative" focus determined by scoring and partitioning heuristics ... rapid focus changes	trace of information acquisition actions (system initiated)
ABEL	"mixed initiative" methodical approach to information acquisition	rationale behind each information acquisition question registered in the dynamically generated plan for information acquisition; user-initiated explanations do not cover for all the system reasoning
NEOMYCIN	"mixed initiative" focus determined essentially by the etiological taxonomy (triggers)	the system reasoning can be explained in abstract terms (strategic principles) and in concrete terms (reasoning tasks carried out); user-initiated explanations cover for all the system reasoning

Explanation structure

The diagnostic plan registers the reasoning that took place in a consultation. For example the rationale behind a question is obtained by ascending the diagnostic plan from the terminal node representing the question. The tree branches represent meta-rules; hence a chain of reasoning would be meaningful if the meta-rules are meaningful. Thus, what needs to be decided is whether the logical bases of the NEOMYCIN meta-rules are sufficiently explicit (why a given condition in the context of some task instantiation would suggest undertaking a particular task?). We would guess that representing strategies in terms of (meta-) rules would suffer from the general problem associated with a pure rule-based representation, namely that exceptions are difficult to encode (Dhar and Pople 1987). Even if the quality of the meta-rules is lacking, NEOMYCIN's explanations attempt to say why a particular reasoning choice was made and not simply that a particular choice has been made. Finally, strategies in NEOMYCIN are expressed in general terms and their instantiations express the application of strategies in specific contexts.

The models of expertise and the human-computer interaction aspects of the analysed systems are summarised in Tables I and II, respectively.

3. Discussion

First we assess the relative 'deepness' of the selected medical expert systems and then we assess their relative qualities of interaction.

3.1 Deepness ranking

For this discussion we are using a comparative definition of deepness which is a generalisation and extension of the definition proposed by Klein and Finin (1987):

A model M can be deeper than a model M' from the following perspectives: (1) the *factual knowledge* in M is more explicit than the factual knowledge in M': (a) M represents domain factual knowledge types which are implicit in M', (b) M gives a richer interpretation to a knowledge type than M' does, and (c) M represents types or aspects of knowledge that are absent in M' even though it would be meaningful to specify such knowledge types and aspects in the context of M'. (2) The *reasoning knowledge* in M is more understandable than the reasoning knowledge in M': (a) M represents reasoning strategies which are implicit in M', (b) M abstracts its reasoning knowledge more than M' does, and (c) the semantics of the reasoning knowledge in M are more explicit than the semantics of the reasoning knowledge in M'. (3) The *solution progression* in M, at intermediate stages, is more understandable than the solution progression in M'

The above definition enables a comparison between systems from a number of deepness perspectives, e.g., from the perspective of reasoning knowledge or the perspective of factual knowledge. As such a system may be considered deeper than another system from one perspective and vice versa from a different perspective. This is also true for the Klein and Finin definition. Deepness is a fuzzy concept and can certainly be viewed from many perspectives in the context of knowledge-based systems like the medical expert systems which exhibit complex organisational structures. Our definition enables a richer multi-dimensional comparison between systems.

In the following paragraphs we discuss our selected systems from these perspectives, concluding NEOMYCIN as the deepest and MYCIN as the least deep system.

NEOMYCIN's model of domain factual knowledge encompasses the *factual knowledge* models of the other systems. This includes an etiological taxonomy, a causal network, triggers, empirical associations between findings and etiologies, disease process knowledge, findings and their interdependencies. Although CASNET has a disease taxonomy this is only used for linking causal pathways to treatment plans and not for focusing the diagnostic process. INTERNIST-I's use of its disease taxonomy in the diagnostic process is rather superficial. The same applies for INTERNIST-I's causal associations between disorders. In addition, INTERNIST-I does not model finding dependencies. ABEL's causality interpretation is the richest among the selected systems. ABEL's causal network at the clinical level is analogous to a disease taxonomy since both structures serve the same purpose, namely focusing of the diagnostic process. ABEL's compiled links are analogous to the trigger links. MDX does not model causality but its model of medical concepts as incorporated in the PATREC auxiliary system is significantly richer than the data models in the other systems.

NEOMYCIN's model of strategic and *reasoning knowledge* is the most explicit among the selected systems. The domain strategic principles are explicit in NEOMYCIN (meta-rules). This is the abstract reasoning knowledge. Instantiations of the reasoning tasks make explicit the application of these domain strategies in actual cases. This is the concrete reasoning knowledge. Statically, the semantics of NEOMYCIN's reasoning are clearer than any of the other systems, due to NEOMYCIN's declarative representation of its reasoning knowledge in terms of tasks and meta-rules. NEOMYCIN has two knowledge-bases, one for the domain factual knowledge and one for the reasoning knowledge. Dynamically, the NEOMYCIN diagnostic plan registers completely the reasoning during a consultation. Again, this is unique in NEOMYCIN. In the other systems the domain strategies are either implicit in some procedural language (CASNET, ABEL and INTERNIST-I), are not represented at all (MDX), or are implicit in mechanistic control structures (MYCIN). ABEL represents its reasoning knowledge entirely in terms of Lisp procedures. During a consultation the system generates information acquisition plans leading to meaningful sequences of questions. Such plans explicate the justification of an information seeking question in terms of the diagnostic hypotheses the question aims to resolve, but the plans do not explicate the strategic principles invoked in selecting the particular questions. MDX's reasoning, in its semi-procedural semi-declarative representation, is more conspicuous than CASNET's and INTERNIST-I's.

In NEOMYCIN, ABEL, CASNET, and MDX the *progressions towards solutions*, at intermediate stages of the consultation, are understandable. This is not so for INTERNIST-I. Only ABEL and CASNET have complex causal hypotheses; NEOMYCIN uses the causal network to hypothesise individual dysfunctional states. The information registered in MYCIN's context tree is understandable. However, the derivation of the parameter values is not necessarily understandable.

Thus, on the basis of our definition, the deepness ranking of the selected systems is: NEOMYCIN, ABEL, CASNET, MDX , INTERNIST-I, MYCIN.

3.2 Human-computer interaction ranking

Apart from MYCIN, all the other systems in our study allow the user to volunteer information. The dialogue structures of CASNET, MDX, ABEL and NEOMYCIN are comparable. ABEL depicts the most organised approach to information acquisition. The diagnostic taxonomies in MDX and NEOMYCIN with their respective triggering mechanisms contribute significantly to the naturalness of the dialogue structures. The CASNET model explicates data dependencies. Its mechanism of assigning weights to states together with the notion of causal hypotheses provide the basis for a meaningful dialogue. From an architectural point of view, however, MDX through PATREC offers a novel approach to intelligent data handling. INTERNIST-I's dialogue structure is superior to MYCIN but inferior to the other systems. INTERNIST-I does not model data dependencies of disease processes (progressions of symptoms in time). Complex hypotheses are not modelled properly and information acquisition is not well planned.

Although the dialogue structures of the selected systems are by and large comparable, this is not so for their respective explanation structures. The ranking from this perspective (which in fact represents the overall human-computer interaction ranking) is: NEOMYCIN, {ABEL, MDX}, CASNET, INTERNIST-I, MYCIN. NEOMYCIN provides the richest explanations and in addition, these explanations are user-initiated and can be offered at any stage during a consultation. MYCIN satisfies the last two criteria as well; however, its explanations simply reproduce rules which hide much of the domain knowledge. NEOMYCIN can explain all its reasoning in both abstract and concrete terms. ABEL and MDX have comparable explanation structures. ABEL explanations are user-initiated but do not cover for all the reasoning that took place in a consultation while MDX explanations (trace of messages exchanged between specialists) are volunteered by the system, and they cover completely the reasoning steps that took place. Neither system is capable of explaining its reasoning in abstract terms and they do not provide any justifications for making a particular decision at a given point in the consultation. CASNET cannot provide any explanations during a consultation, but it can explain its conclusions on user request at the end of a consultation. These explanations do not reveal the actual system reasoning. During a consultation INTERNIST-I displays its conclusions, the current problem area, and the information acquisition strategy used to resolve the current problem. CASNET's explanations, although only available at the end of a consultation, are more useful than INTERNIST-I's volunteered explanation trace.

In the literature of medical expert systems a deep system tends to imply a causal system. If we were to rank the selected systems from the perspective of how strongly they model and reason with causality then this ranking would have been: ABEL, CASNET, NEOMYCIN, INTERNIST-I, {MDX, MYCIN}. Our deepness ranking correlates closer to the human-computer interaction ranking than the above causality ranking. In a medical expert system causality is probably a necessary but not a sufficient condition for deepness. The concept of causality is not in fact absolute; one can always define more detailed description of some causal phenomenon. The fact that a model represents some kind of causality should not automatically qualify it as a deep model; a causal model may still be shallow in relative terms. For example, ABEL has three causal networks; the topmost network (clinical level) certainly does not embody detailed knowledge.

4. Conclusion

Expert systems in general and medical expert systems in particular have not quite reached the high levels of performance exhibited by the human experts; they can often deal very well with instances of common problems but their performance degrades drastically when dealing with a difficult case which doesn't quite fit the norms. Expert systems do not exhibit the flexibility that characterizes human expertise and they are not in a position to recognize when a problem does not belong to their particular area of expertise. In addition, expert systems cannot explain their reasoning and conclusions in meaningful ways and often cannot converse intelligently with the user; since expert systems are strongly interactive systems these are serious deficiencies.

Realization of the above problems with the technology resulted in the generation of the so-called *deep systems*. Why expert systems have not quite met their expectations is attributed to their being *shallow*, i.e., they lack the depth of understanding that human experts have. A deep expert system, or second generation expert system, should exhibit the problem solving flexibility of a human expert and thus reach truly high levels of performance.

Deepness has an appealing intuitive meaning but it is not easy to assign clear semantics to it. This is evident from the numerous definitions which have been put forward for a deep expert system. In the context of medical expert systems the most relevant definitions involve causality and temporal reasoning. As we have seen above causality is a relative concept and the ability to reason at different causal levels is more useful than the ability to have a single causal level (cf. CASNET and ABEL); of course, the more abstract causal levels are just like the empirical associations of shallow systems. Problem solving flexibility derives from the ability to apply more than one reasoning method and to switch from one method to another as the particular problem case dictates (cf. NEOMYCIN which can either apply a top to bottom refinement strategy using the triggers and etiological taxonomy or a left to right differential diagnostic strategy using the causal network or a combination of these). Solving an easy problem case by applying the method for difficult cases is neither efficient nor effective. More importantly, this approach cannot yield a human-computer interaction that is acceptable. A uniform reasoning method with its corresponding knowledge structures is not sufficient. This is where current proposals for deep expert systems go wrong, for example Davis' proposal that a deep system must reason from first principles, or reason from structure and function (Davis 1983). Such proposals aim for a single reasoning method which could solve any problem case. The consequence of this is that the chosen method (and representation formalism) does not emulate human expertise (Keravnou 1986), thus defeating the original motivation for having a deep system. The coexistence of different reasoning methods requires the integration of multiple representations for the domain knowledge and the ability to move from one representation to another. Instances of common problems can be dealt with in one way and instances of difficult problems in another; only then is the knowledge put into *effective* use.

It is generally accepted that shallow systems are only capable of matching problem instances against predetermined patterns which subsequently lead to decisions. These patterns have been compiled, presumably by the domain experts, through experience and deal with the majority of cases but not every single case. If a particular problem fits such predetermined patterns then its solution can be derived very quickly. Deep understanding of a domain on the other hand implies knowledge of how the various domain entities interact with each other so that a particular instantiation at a certain point of the problem space can be propagated through the space. Such understanding enables the handling of cases that do not fit the compiled patterns. However, trying to solve a case that fits the compiled patterns in this way displays naivety rather than expertise. A deep system must combine a deep domain understanding with the experiential domain knowledge. Chandrasekaran and Mittal (1982) bring out this issue very succinctly, drawing from the domain of cholestasis diagnosis. They suggest that an expert system must explicate the conceptual structure of the domain knowledge; it is this structure, acquired through experience, which enables the effective use of that knowledge.

Our definition of deepness relates closely to what Chandrasekaran and Mittal are proposing. We propose that a deep expert system must properly articulate human expertise, i.e., explicate the domain structure and the reasoning methods used by the expert. This is in line with Chandrasekaran's generic task methodology (Chandrasekaran 1986). The semantics of the generic reasoning methods must be explicit in the system; these semantics explain if and when one switches from one reasoning strategy to another, in context. Expert problem solving is usually characterized with unknown or imprecise information. Much of an expert's expertise (especially in diagnosis) is in acquiring new information and reasoning in the absence, or acquisition, of previously unknown information. An expert's reasoning is therefore nonmonotonic; new strategic choices are made as the case specific information grows. A deep expert system must be capable of such non-monotonic reasoning at the strategic level if it is so to exhibit the flexibility of a human expert and also yield an acceptable human-computer interaction.

A deep system must therefore deal effectively with both common and uncommon instances of problem cases, i.e., exhibit high problem solving flexibility, and must exhibit the expectations of its users regarding meaningful explanations and a comprehensible dialogue. The implication of these requirements is that the system must provide a smooth amalgamation of all the reasoning methods (generic tasks) and knowledge structures used by the human experts. The rationale for applying each of the generic tasks in an actual context must be made explicit to allow the system to reason effectively, in a non-monotonic fashion, at the strategic level. This essentially means that a deep system must be able to plan its reasoning strategy for a specific case and modify the plan when new case specific information violates the rationale underlying the application of a particular plan step.

We conclude by giving our working absolute definition for a deep system: a deep system is one that adequately explicates the models of its domain factual and reasoning knowledge and can reason non-monotonically at the strategic level.

Acknowledgements. We would like to thank Professors J.A. Campbell (University College London) and L. Johnson (Brunel University) for their constructive comments on earlier drafts of this paper.

References

Chandrasekaran, B.: 1986, Generic tasks in knowledge-based reasoning: High level building blocks for expert system design. *IEEE Expert*, Fall 1986, 23–30.

Chandrasekaran, B., Gomez, F., Mittal, S., and Smith, J.: 1979, An approach to medical diagnosis based on conceptual structures. *Proc. IJCAI-79*, 134–142.

Chandrasekaran, B. and Mittal, S.: 1982, Deep versus compiled knowledge approaches to diagnostic problem-solving. *Proc. AAAI-82*, 349–354.

Chandrasekaran, B. and Mittal, S.: 1983, Conceptual representation of medical knowledge for diagnosis by computer: MDX and related systems. *Advances in Computers* **22**, 217–293.

Clancey, W.J.: 1986, From Guidon to Neomycin to Heracles in twenty short lessons: ORN final report 1979–1985. *AI Magazine*, August 1986, 40–60.

Clancey, W.J. and Letsinger, R.: 1981, Neomycin: Reconfiguring a rule-based expert system for application to teaching. *Proc. IJCAI-81*, 829–836.

Clancey, W. and Shortliffe, E.R. (eds.): 1984, *Readings in Medical Artificial Intelligence: The First Decade*. Addison-Wesley, Reading, MA.

Davis, R.: 1983, Reasoning from first principles in electronic troubleshooting. *Int. J. Man-Machine Studies* **19**, 403–423.

Dhar, V. and Pople H.E.: 1987, Rule-based versus structure-based models for explaining and generating expert behaviour. *Communications of the ACM* **30**, 542–555.

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

Keravnou, E.T.: 1986, Automated reasoning in fault diagnosis and verification. *Int. J. Systems Research and Information Science* **1**, 269–286.

Keravnou, E.T. and Johnson, L.: 1986, *Competent Expert Systems: A Case Study in Fault Diagnosis*. Kogan-Page, London.

Klein, D. and Finin, T.: 1987, What's in a deep model: A characterization of knowledge depth in intelligent safety systems. *Proc. IJCAI-87*, 559–562.

Miller, R.A., Pople, H.E., and Myers, J.D.: 1982, Internist-I: An experimental computer-based diagnostic consultant for general internal medicine. *New England J. of Medicine* **307**, 468–476.

Patil, R.S.: 1981, Causal representation of patient illness for electrolyte and acid-base diagnosis. *MIT/LCS/TR-267*.

Price, C.J. and Lee, M.: 1988, Deep knowledge tutorial and bibliography. *Alvey Report IKBS3/26/048*.

Shortliffe, E.H.: 1976, *Computer-Based Medical Consultations: MYCIN*. American Elsevier Publishing Co., New York.

Shortliffe, E.H., Buchanan, B.G., and Feigenbaum, E.A.: 1979, Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proc. IEEE* 67, 1207–1224.

Weiss, S.M.: 1974, *A System for Model-Based Computer-Aided Diagnosis and Therapy*. Ph.D. Thesis, Computers in Biomedicine, Department of Computer Science, Rutgers University, CBM-TR-27-Thesis.

Young, D.W.: 1982, A survey of decision aids for clinicians. *British Medical Journal* **285**, 1332–1336.

Authors' address

E.T. Keravnou and J. Washbrook, Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK