Causes to clauses: managing assumptions in qualitative medical diagnosis

Keith Dowsing
University of Oregon, Eugene, Oregon 97403, USA

Jeff Shrager
Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

The ABD project demonstrates the applicability of formal techniques of assumption-based reasoning to qualitative diagnosis, a task complicated by many types of assumptions. Our primary goal is to connect the medical-knowledge level to the logical level by automating the conversion of high-level diagnostic concepts such as 'symptoms' and 'faults' into the constructs of the Assumption-Based Truth Maintenance System (ATMS). Due to the simplicity of the ATMS logic, this compilation requires the formalization of many of the implicit assumptions of diagnostic reasoning. When the proper assumptions are included, the results of interpretation construction will correspond to all diagnostic explanations consistent with the fault and symptom information provided by the physician.

Key Words: qualitative reasoning, truth maintenance, assumption management, nonmonotonic reasoning.

1. INTRODUCTION

This paper presents a method of qualitative medical diagnosis based on assumption management. The Assumption-Based Diagnostician (ABD) links the knowledge level to the logical level by compiling medical concepts into propositional logic, where ATMS interpretation construction is used to generate diagnostic explanations.

The motivation for using an assumption-based approach arises from three important processes of medical diagnosis:

1. Causal Abduction - If the domain knowledge is encoded in rules having causes as antecedents and effects as consequents, then the process of constructing an explanation of symptoms in terms of causes is backward reasoning (abduction) over those causal rules.

2. Theory Management - A typical diagnostic cycle consists of observation followed by the formation of explanatory theories which are then tested and accepted, rejected, or revised.

3. Qualitative Reasoning - Medical experts explain causality mainly in terms of qualitative derivatives and ordinal relationships between quantities. For example, a common description of the causal pathway through which sodium intake affects blood pressure is:

- Increased sodium intake results in a stronger concentration of plasma sodium, which raises the oncotic pressure between the interstitial (between cell) spaces and the plasma -- causing the osmotic of water into the blood stream. The ensuing elevation of blood volume then initiates a rise in cardiac blood output, which raises the blood pressure.

(Abstracted from 11.)

All of these processes involve a great deal of nonmonotonic reasoning. In abduction and theory formation, causal explanations represent defeasible assumptions, because in the absence of closed-world assumptions (e.g., 'High salt intake and nervous tension are the only possible causes of high blood pressure.') such explanations can not be logically deduced from the observed effects. Furthermore, the local nature of constraint propagation and the nondeterminism of qualitative arithmetic leads to ambiguity in qualitative reasoning -- thereby requiring auxiliary assumptions, such as the confluence heuristics de Kleer and Brown.

Qualitative reasoning therefore adds yet another degree of nondeterminism to the inherently nonmonotonic process of diagnosis, making assumption management a vital aspect of qualitative diagnosis.

The ABD system converts high-level causal relationships between faults and symptoms into an ATMS causal network. Then, when provided with a set of symptom premises, the ATMS performs interpretation construction, which returns all possible collections of fault and auxiliary assumptions capable of deriving those symptoms via the causal justification structure. These possible diagnoses could then be evaluated relative to criteria such as the likelihood of the fault assumptions alone or in combination with others, or the feasibility of the auxiliary assumptions.

For instance, when provided with the symptom of increased arterial pressure, ABD might generate the multiple-fault hypothesis: 'elevated blood volume and decreased peripheral resistance'. Since increased blood

© Computational Mechanics Publications 1988

192 Artificial Intelligence in Engineering, 1988, Vol. 3, No. 4
Managing assumptions in qualitative medical diagnosis: K. Downing and J. Shraer

volume alone would cause arterial pressure to rise, while a peripheral resistance drop would cause it to fall. ABD fortifies this diagnosis with the auxiliary assumption that the effect of blood volume overrides that of peripheral resistance. Consequently, arterial pressure rises in spite of the conflicting influences. A physician could recognize this diagnosis as the initial stage of volume-loading hypertension and then use knowledge of the patient’s recent fluid-intake to confirm or deny this explanation.

The first part of this paper introduces Causally-Bounded Diagnosis (CBD) as a methodology for carving a diagnostic environment out of a complex qualitative model. Then, we show how the dichotomy between symptoms and faults in the CBD network enables the translation of this network into an ATMS causal structure. This conversion makes ATMS interpretations/explanations correspond to diagnoses in the CBD model. Unfortunately, the CBD analysis results in many more explanations that are reasonable. In the last part of this paper we show how ambiguity can be constrained by introducing explicit meta-assumptions, such as the single-fault assumption. In addition, meta-assumptions define an abstraction hierarchy of explanations for qualitative assumption-based diagnosis.

2. FORMALIZING DIAGNOSTIC PROBLEMS IN BIOLOGICAL DOMAINS

In many electrical and mechanical domains, diagnosis can be aptly described as ‘reasoning from unusual behaviors (symptoms) to structural abnormalities (faults)’.

The pairing of faults with structural defects seems accurate in these domains, because many faults involve major structural aberrations such as shorted resistors, disconnected wires, or dead motors. Fortunately for living organisms, most ailments do not involve major structural damage; totally severed arteries, non-functioning organs and similar traumas occur less frequently than shifts of equilibrium conditions to ‘illness’ states. Hence, medical diagnosis attempts to explain observed changes in certain symptomatic parameters by changes in other parameters that could cause those symptoms. The ‘fault’ is often some primary causal parameter whose proper modification (via therapy) will relieve the problem.

Still, it is often difficult to discern primary from secondary causal factors, especially in feedback environments (such as the human body). Furthermore, a deeper physiological analysis often reveals a cause for a seemingly independent causal agent. In short, the tenuous dichotomy between symptoms and faults has no absolute basis; it only exists relative to factors such as the available therapies, the region of the body under diagnosis, the granularity of diagnostic reasoning, and the relevant forms of causal interaction.

For example, a cardiologist, who deals primarily with hydraulic and neuro-electrical forms of causation, would view a clogged artery as a fault that contributes to the symptom of high blood pressure: on the other hand, a dietician, whose main concern is the chemical composition of the blood, would describe the same blocked artery as a symptom of an excessive cholesterol concentration.

Any formalization of diagnostic reasoning requires some commitment to a specific symptom/fault division.

In previous work this was achieved by introducing a ‘Causally-Bounded Diagnosis’ (CBD). The CBD relativizes symptoms and faults to a ‘causal border’ defined by some mechanism of causal interaction. For instance, relative to a hydraulic causal mechanism, clogged arteries are faults, while elevated blood pressure is a symptom, since only the latter results from hydraulic activity and is a true effect from the hydraulic standpoint (Fig. 1).

The CBD involves two steps:

1. Causal Circumscription – Given a network consisting of different types of causal relationships between system parameters, choose one mechanism of causal interaction (e.g., hydraulic, electrical, or kinematic). Use this ‘relevant’ mechanism to circumscribe some portion of the total causal network so that all parameters within the circumscription are involved in at least one causal relationship of the selected type. Diagnosis will take place only within the circumscription; hence diagnosis will involve only the circumscribed parameters and the selected causal relationships between.

2. Symptom-Fault Relativization – Define fault parameters as those parameters not causally influenced by other circumscribed parameters. Faults are thus independent and static relative to the circumscription since their only influences come from uncircumscribed parameters or ‘irrelevant’ causal mechanisms. The dependent or symptomatic parameters are those causally affected by at least one circumscribed parameter.

The CBD in Fig. 1 includes the causal pathway from sodium intake to blood-pressure change. The arc labels ‘+’ and ‘−’ designate direct and inverse causal influences, respectively, and are further annotated by the type of causal interaction. Choosing the hydraulic activities of the circulatory loop as the relevant causal mechanism yields the indicated circumscription (dashed box). Relativization designates total blood volume (BV), venous compliance (VC), and arterial resistance (AR) as fault parameters, while cardiac output (CO) and arterial pressure (AP) represent symptomatic parameters. Hence the single-fault hypotheses that explain ΔAP+, i.e., those causally-related parameters which lie on the internal border of the circumscription, are ΔBV+, ΔVC−, or ΔAR+.

Artificial Intelligence in Engineering, 1988, Vol. 3, No. 4 193
3. THE ATMS AS AN EXPLANATION GENERATOR

Since Assumption-Based Diagnosis relies upon ATMS-based diagnosis, it is important to understand ATMS interpretation construction. This section summarizes the relevant aspects of the ATMS. The next section describes our application of ATMS reasoning in Causally-Bounded Diagnosis.

The ATMS is a database system which can be used by a problem-solver to maintain information about the state of the problem solution. Unlike typical truth maintenance systems the ATMS can explore all possible consistent solutions. It does so by maintaining the assumptions that underlie each conclusion.

The ATMS stores data in nodes, which are connected by justifications. The justification records how a node was derived from other nodes. Three types of nodes are important for this discussion: Premises are true in all contexts — they require no justifications. Assumptions are self-justifying, and are therefore defeasible premises. Conclusions are justified by other premises, assumptions, or conclusions.*

Nodes may contradict one-another. A collection of non-contradictory assumptions constitutes an environment. A context denotes the deductive closure (via justifications) of an environment. A nogood is a minimal set of contradictory assumptions — that is, one with no contradictory subsets. An interpretation is a maximal environment — that is, one with no consistent supersets.

Thus, each interpretation has a single context which is the set of nodes derivable from the interpretation’s assumptions.

The following simple example illustrates these concepts:

\[ \begin{align*}
A &\Rightarrow e \\
e, B &\Rightarrow f \\
\text{nogoods: } (A,C), (C,D) \\
C &\Rightarrow e \\
\text{interpretations: } (A,B,D), (B,C) \\
D &\Rightarrow d \\
\text{contexts: } (A,B,D,e,f,d), (B,C,e) \\
c, d &\Rightarrow \perp \\
e, c &\Rightarrow \perp
\end{align*} \]

An ‘explanation’ for a collection of nodes is an interpretation whose context contains those nodes. Hence,

\[ \text{explanation}(f) = \{(A, B, D)\}^* \]

while

\[ \text{explanation}(f,c) = \{ \} \]

Thus, an ATMS explanation is a maximal consistent set of assumptions capable of deriving the desired nodes through the justifications. This mirrors Poole’s default-logic definition of explanation through extension membership. Like Poole, we can exploit this definition of explanations to characterize diagnoses.

4. CAUSALLY-BOUNDED DIAGNOSIS WITH THE ATMS

A Causally-Bounded Diagnosis problem can utilize the ATMS explanation capabilities once the CB network has been translated into the proper collection of ATMS contexts. We use a simple causality model to represent the ATMS system. The image shows a simple example of the model.

Causally-Bounded systems require a complex network of assumptions, conclusions, nogoods, and justifications. Medical diagnosis requires these models because the interactions occur quite frequently in self-regulating systems like the human body, the sympathetic nervous system increases arterial and venous compliance and decreases heart rate. This section illustrates a progression of models that appropriately capture the semantics of multiple interacting faults.

Since fault parameters have no causes within the circumscription, they parallel assumptions, which represent ‘deducitively independent’ datum. Similarly, since symptoms are causally affected by other circumscribed parameters, they are appropriately characterized by conclusions, which are justified by other nodes.

In a qualitative first-order system, the relevant property of each parameter is the sign of its derivative. Hence, three mutually-exclusive and exhaustive sets: \( dX = 0 \), \( dX = - \), and \( dX \neq 0 \). Components represent all possible state of the parameter \( X \). Justifications between the modes of different parameters reflect causal influences. For instance, if \( X \) directly influences \( Y \), then the justifications:

\[ dX + \Rightarrow dY +, \quad dX - \Rightarrow dY - \]

capture that relationship. Similarly, if \( X \) inversely influences \( Y \), then these justifications are added:

\[ dX + \Rightarrow dY -, \quad dX - \Rightarrow dY + \]

The mode in which \( X \) is constant (\( dX = 0 \)) never serves as an

* Notice that the ‘maximal’ interpretation forces the inclusion of the irrelevant assumption \( D \) in the explanation of \( f \). The ATMS defaults on many assumptions as possible during interpretation construction — yielding many overspecific explanations. We are exploring an approach to this problem in current research.

194 Artificial Intelligence in Engineering, 1988, Vol. 3, No. 4
Fig. 3. The model of opposing influences upon CO under the opposition-cancellation constraint

antecedent in any justification. This reflects the view that an unchanging parameter has no causal influence. However, due to the maximality of the ATMS interpretations, these modes will appear in explanations whenever the modes are consistent.

The ATMS structure of Fig. 2 encodes a part of the CBD situation of Fig. 1. When \( \delta AP^+ \) (elevated arterial pressure) is taken as a premise, interpretation construction yields:

\[
\begin{align*}
\text{explanation}(\{\delta CO\}) &= (\delta BV^+, \delta VC^-) \\
&= (\delta BV^+, \delta VC^-) \\
&= (\delta BV^+, \delta VC^-)
\end{align*}
\]

Hence, the explanation of high arterial pressure (\( \delta AP^+ \)) by a rise in blood volume (\( \delta BV^+ \)) includes the assumption that venous compliance either holds steady (\( \delta VC^- \)) or decreases (\( \delta VC^- \)). If both blood volume and venous compliance increase, then the veins tend to hold more of the additional blood, thereby delaying its return to the heart. Consequently, increased venous compliance counteracts the effects of increased blood volume upon cardiac output and arterial pressure. This opposition of influences is evident in Fig. 2, where (\( \delta BV^+ \), \( \delta VC^- \)) constitutes a nogood, since the two assumptions derive contradictory values for \( \delta CO \).

The ATMS causal network of Fig. 2 obviously cannot explain symptoms in terms of opposing faults, since any combination of opposing faults will derive a contradiction. Since the undesirable contradiction occurs at the cardiac-output (\( \delta CO \)) level, we will restrict our focus to \( \delta CO \), rather than \( \delta AP \), for the remainder of this discussion. In Fig. 3, the introduction of an intermediate ‘influence’ level partially alleviates this problem by allowing opposing influences to exist without contradiction. With this addition, the observation level represents the resultant change in the parameter (i.e., the qualitative sum of the influences). As a temporary simplification, the network of Fig. 3 manifests the Opposition-Cancellation Constraint:

Any number of opposing influences upon the same parameter resolve to no net change in that parameter, i.e.,

\[
(\text{Influence } \delta CO^+ \text{, Influence } \delta CO^-) \rightarrow \text{Observation } \delta CO
\]

Although the influence level permits opposing influences, it merely delays the derivation of contradictions until the \( \delta CO \) observation level. So, (\( \delta BV^+ \), \( \delta VC^- \)) is once again a nogood, because this fault pair derives all three symptom modes at the observation level. This problem occurs because a negative (or positive) influence derives a negative (or positive) observation even when an opposing positive (or negative) influence exists. Under the Opposition-Cancellation Constraint, these should resolve to no change (\( OBS \delta CO \)), hence an influence should only derive its equivalent observation in the absence of any opposing influences. ATMS assumptions are needed to account for this absence. The network of Fig. 4 achieves the desired behavior through two ‘non-influence’ assumptions.

These are added to an explanation when and only when no influences of the given type (+ or −) exist. With these refinements, the model sanctions (\( \delta BV^+ \), \( \delta VC^- \)) as a viable explanation, but only for \( \delta CO \). Due to the Opposition-Cancellation Constraint, opposing influences cannot resolve in any other way without additional assumptions.

Interpretation construction over the CBD model of Fig. 4 yields:

\[
\begin{align*}
\text{explanation}(\{\delta CO\}) &= (\delta BV^+, \delta VC^+) \\
&= (\delta BV^-, \delta VC^-)
\end{align*}
\]

and

\[
\begin{align*}
\text{explanation}(\{\delta CO\}) &= (\delta BV^+, \delta VC^-) \\
&= (\delta BV^-, \delta VC^-)
\end{align*}
\]

Our final enhancement to the ATMS Causally-Bounded Diagnostic model removes the Opposition-Cancellation Assumption and accounts for the ambiguity of qualitative arithmetic by supplementing the network of Fig. 4 with three assumptions concerning the ‘resolvents’ of opposing influences (Fig. 5). Now the opposing influences can resolve to a positive, negative, or zero observation, but only in combination with the appropriate resolvent assumption. Hence, in the causal network of Fig. 5, in which the resolvents RES\( \delta CO \), RES\( \delta CO^- \), and RES\( \delta CO^+ \) have been added, (\( \delta BV^+ \), \( \delta VC^- \)) can explain \( \delta CO \) when RES\( \delta CO^- \) is included in the explanation.

The entire set of explanations for elevated cardiac output is now:

![Fig. 4. The model of opposing influences upon CO including noninfluence assumptions](image-url)
Fig. 5. Complete ATMS model of opposing influences on CO including resolvents which relax the opposition-cancellation constraint

\[
\begin{align*}
(\text{[ECO +]}) \\
(\text{[BV+}, \text{VCO}, \text{-INF OCO}]) \\
(\text{[BV+}, \text{VIC-}, \text{-INF OCO}]) \\
= (\text{[BV+}, \text{VIC-}, \text{-INF OCO}]) \\
(\text{[BV+}, \text{VIC+}, \text{RES OCO}]) \\
(\text{[BV-}, \text{VIC-}, \text{RES OCO}])
\end{align*}
\]

To summarize, the stages in the evolution of a representation for qualitative causal knowledge were (a) the simple causal model, (b) the inclusion of the influence level, (c) the introduction of the non-influence assumptions, and (d) the addition of the resolvent assumptions. The benefits of the final layered representation are that:

1. it allows fault parameters to exert opposing symptomatic influences without deriving a contradiction.
2. it explicates some of the deep implicit assumptions of qualitative causal reasoning and incorporates them in diagnostic explanations.
3. it modularizes the knowledge structures for symptoms and faults and restricts their interactions to the influence level.
4. it simplifies the incremental creation and refinement of the CBD environment (as a result of the aforementioned modularity).

In this model, explanations contain fault assumptions along with the auxiliary assumptions required to derive the symptom(s). These auxiliary assumptions represent intuitively subtle, yet logically necessary factors used to derive symptoms from faults. For instance, \(\text{BV+} \) cannot explain \(\text{OCO}+\) without the auxiliary assumption \(\text{-INF OCO}+\). These assumptions add another aspect to diagnostic decision making.

Within a given causal circumscriptive, assumptions and causally-derived data play different roles in the evaluation of diagnostic hypotheses (i.e., collections of fault assumptions). Assumptions are the debatable components of a diagnosis. Auxiliary assumptions, such as the resolvent assumptions in the previous example, indicate the subtle claims the diagnostician must believe to confidently accept the fault assumptions. Causally-derived information is what he must observe to empirically verify the fault predictions. Thus, derived information connects hypotheses to the physical world by indicating the most potentially-informative test parameters. Assumptions, on the other hand, can link diagnoses to external (i.e., outside the circumscriptive) causal, statistical, or practical knowledge - all of which can help constrain reasoning within the causal circumscriptive.

For instance, the explanation of increased arterial pressure by the fault combination of high blood volume and increased venous compliance will require the auxiliary assumption \(\text{RES AP+}\), which indicates that \(\text{BV}\)'s positive influence upon \(\text{AP}\) overrules \(\text{VC}\)'s negative effect. This diagnosis would garner support from external causal knowledge stating that (a) the nervous system normally regulates elevated blood volume by increasing venous compliance, and (b) this feedback has finite gain, so \(\text{AP}\) will rise not nearly as much as when changes in \(\text{BV}\) go unchecked. On the other hand, knowledge that the patient has just finished exercising would refute this hypothesis, since physical exertion generally raises \(\text{BV}\) and \(\text{AP}\) but decreases \(\text{VC}\). In either case, the cerebral information comes from outside the causal explanation. Hence, assumptions serve as binding sites for peripheral knowledge. This helps evaluate diagnostic hypotheses prior to experimentation.

Those explanations that survive filtering by external knowledge can be expanded into their respective ATMS contexts (i.e., the expected observations derivable from a set of faults) which may guide laboratory or clinical evaluations.

Unfortunately, this generate-and-test interaction between interpretation construction and knowledge-directed pruning of explanations results in an 'interpretation bulge': the intermediate set of interpretations is much larger than the set after filtering (i.e., knowledge-directed pruning) - because all possible interpretations consistent with the basic causal circumscriptive are created.

For instance, consider a diagnostic environment in which \(n\) fault parameters influence the symptom \(Y\). Assume, \(\text{BV+}, \text{VIC+}\), and assume that one of the \(n\) fault parameters directly influences \(Y\). The \(\text{BV+}+\) along with \(\text{VIC+}\) with any combination of other modes for the remaining \(n-1\) parameters will explain \(\text{BV+}\) (when the appropriate non-influence and resolvent assumptions are included in the explanation). So, any interpretation in which at least one of the \(n\) faults positively influences \(Y\) will explain \(\text{BV+}\). There are \(3^n-2^n\) such interpretations. Consequently, when \(n=10\), interpretation construction yields over 50,000 possible explanations.

One promising remedy involves the compilation of external or meta-constraints - such as the assumption that \(\text{BV}\) is stronger than \(\text{VC}\) in its effect upon \(\text{AP}\) - into the basic causal network. By providing additional constraint on interpretation construction, this compilation effectively moves tests into the generator and reduces the intermediate interpretation bulge.

5. META-ASSUMPTIONS

The ambiguities of qualitative arithmetic and the simplicity of propositional logic lead to many implausible but logically possible, diagnostic explanations. One example is the hypothesis that all independent parameters are faulted. To prune the set of possible explanations

In a set of four nogoods, each of which contains the influence-dominance assumption $\text{Dominates}(BV, VC, CO)$:

$\text{Explains}(CO) \neq \text{Influence}(CO)$

By declaring $\text{Dominates}(BV, VC, CO)$ true and thereby forcing every interpretation to contain it, the interpretation constructor is restrained by the new nogoods and generates a smaller set of explanations for $\text{CO}+$.

The addition of the Single-Fault Assumption (SFA) results in four more nogoods:

$\text{SFA,} \delta BV+, \delta VC+$

These further restrict interpretation construction to two explanations:

The meta-assumptions characterize a hierarchical approach to diagnosis wherein higher levels of abstraction contain more meta-assumptions (and hence are more highly constrained by nogoods) then lower regions. This meshes with the view of abstraction as an assumption-making process in which higher levels 'assume away' the details of lower levels by positing fixed values of, and fixed relationships among,
Managing assumptions in qualitative medical diagnosis: K. Downing and J. Shrarer

'abstracted' system parameters. Hence, the most 'abstract' levels actually admit the fewest plausible explanations. These tiers are abstract precisely because assumptions have frozen many details, leaving fewer to consider.

Thus, ABD provides an operational logical formalism for a hierarchy of diagnostic assumptions similar to Davis' k-or-ordered categories of failure. However, Davis' hierarchy organizes multiple representations as defined by alternate mechanisms of causal interaction; but Causally-Bounded Diagnosis defines the basic diagnostic environment relative to the causal circumscription carved out by a single causal mechanism. Meta-assumptions then constrain behavior within the circumscription. Hence, our assumptions restrict reasoning within causal levels while Davis' assumptions predominantly govern shifts between causal representations. The fundamental difference between our and Davis' is that ABD meta-assumptions arise from the need to methodically control the ambiguities of qualitative diagnosis. So while Davis' assumptions enable the expansion of quantitative diagnosis, one for each representation, ABD portraits the myriad assumptions necessary to reason within a single qualitative representation.

The encoding of meta-assumptions as nogoods also provides a way to measure the reduction in the explanation space - i.e., the bias introduced by the meta-assumptions. For instance, in a system with n fault parameters, the single fault assumption reduces the explanation space from $2^n$ (3 possible modes for each fault parameter) to $n+1$ (in possible faults plus the no-fault case).

In a more complicated case, assume that symptomatic parameter Y has a zero derivative. Given that the fault parameters $X_i$ and $X_j$ influence Y along with k other fault parameters, we can calculate the bias of an influence-domination meta-assumption that states $X_i$ dominates $X_j$ relative to Y. This bias is the number of fault-mode combinations that would satisfy the nogood schema which has already been partially instantiated with $\delta Y$. Of the $3^k$ possible combinations of $\delta X_i$ and $\delta X_j$ values, two are relevant to the meta-assumption -- namely, the two cases where $\delta X_i$ and $\delta X_j$ oppose one-another in influencing Y. Furthermore, to satisfy the no good schema, (as equation (10)) none of the other k fault parameters can exert a Y influence equal to that of $\delta X_j$. Hence, one mode for each of the k parameters must be dismissed, leaving $2^k$ combinations. The remaining $n-k-2$ fault parameters can assume any mode, since they have no effect on Y. Consequently, the meta-assumptions prune $2^{k+1}$ explanations. In the maximal case, every fault parameter affects Y; thus $k=n-2$, and $2^{n-2}$ explanations are sacrificed. In the minimal case, k = 0 and 2 $3^n$ explanations are pruned.

Thus, the meta-assumption has a greater effect when fewer additional faults influence the specified symptom, simply because the truth of the nogood schema is contingent on fewer parameter-mode assumptions.

To summarize, by restricting interpretation construction, meta-assumptions compensate for the extreme generality of the basic ATMS causal network, which otherwise leads to an overabundance of possible explanations for given symptoms. These defeasible, nonmonotonic meta-constraints define a hierarchical approach to diagnosis in which the higher abstraction levels embody stricter notions of 'plausibility' than do the lower, less-constrained levels. The encoding of meta-assumptions as nogoods eases movement between abstraction levels and helps to quantify the biases imposed on interpretation construction. In a nutshell, meta-assumption compilation moves much of explanation testing into the generator - thereby alleviating the intermediate-interpretation bulge and making assumption-based diagnosis more manageable.

6. THE ASSUMPTION-BASED DIAGNOSTICIAN

The Assumption-Based Diagnostician (ABD)* works as follows:

1. The physician enters the symptomatic and fault parameters along with the qualitative influences that relate them. This defines the diagnostic environment.

2. ABD compiles the parameters and influences into an ATMS causal network by introducing the ATMS assumptions, nodes, justifications, and contradictions that define the logical semantics of the high-level concepts.

3. At any time after step 2, the physician can enter meta-assumptions, which ABD compiles into nogoods. Meta-assumptions can be deactivated and reactivated when desired.

4. The physician enters specific symptomatic observations to define a particular diagnostic problem.

5. ABD biases the causal network by converting to premises those nodes corresponding to the physician’s symptomatic observations.

6. ABD runs ATMS interpretation construction over the biased causal network, generating all possible explanations in terms of fault, auxiliary and meta-assumptions.

Steps 3 through 6 may be repeated as many times as desired with different observations and/or meta-assumptions. This allows the physician to fully explore the explanation space while concentrating only on the pivotal assumptions underlying his decision-making.

7. DISCUSSION

De Kleer and Brown* detail the role of implicit assumptions in modelling physical systems. They indicate that both novices and experts employ assumptions to direct reasoning. However, experts have explicated their assumptions and thus recognize the limitations of their models while also understanding how to transform those models by adjusting certain crucial assumptions. Assumption management is thus an integral part of expert reasoning. In addition, the non-monotonicity of diagnosis and the ambiguity of qualitative reasoning suggest that assumption management should play a major part in automated diagnosis.

* This directly contrasts with an ignorance-based approach in which abstract reasoning ignores certain parameters under the assumption that none of their possible values or relationships could affect high-level reasoning. Automated induction systems typically create this type of abstraction in their search for the salient features of training examples.

* ADB was developed in Xerox Commonlisp and runs in conjunction with de Kleer's ATMS and the Xerox NoteCards System.

198 Artificial Intelligence in Engineering, 1988, Vol. 3, No. 4
The ABD project formalizes some of the tacit assumptions of medical diagnosis. The ATMS was developed to manage assumptions made during problem solving, especially in situations demanding multiple solutions. De Kleer and Williams have demonstrated the ATMS's utility for handling the nonmonotonicity of hypotheses generation and testing during diagnosis; but they only discuss ATMS assumptions representing beliefs that specific components are functioning properly. Many other types of assumptions participate in the diagnostic process but currently lack appropriate logical semantics. ABD formalizes some of these assumptions and illustrates their participation in diagnostic explanation. These new assumptions fill many of the gaps between the diagnostic-knowledge level and the logical level, and methodologies for managing these assumptions characterize levels of diagnostic expertise.

The nature of interpretation construction poses knowledge-representation problems. It is important that the logical semantics of the additional assumptions mirrors their knowledge-level semantics. If so, the interpretations will constitute meaningful explanations rather than collections of ad-hoc assumptions having little relevance for diagnostic reasoning. ABD grounds de Kleer and Brown's theories of implicit assumptions in the ATMS — thereby bringing together two lines of research central to qualitative reasoning.

ACKNOWLEDGEMENTS

Thanks to Johan de Kleer for his advice on using the ATMS on qualitative physics, and for commenting drafts of this paper. Nils Peterson helped formulate the cardiovascular models, and Sally Douglas provided general advice and support.

REFERENCES

1 Campbell, K. The Physical Basis of the Problem Environment in Cardiovascular Diagnosis, January 1983, unpublished
3 de Kleer, J. An Assumption-Based TMS, Artificial Intelligence, 1986a, 28, 127-161
4 de Kleer, J. Extending the ATMS, Artificial Intelligence, 1986b, 28, 163-196
8 de Kleer, J. and Williams, B. C. Diagnosing Multiple Faults, Artificial Intelligence, 1987, 32(1), 97-130
10 Forbus, K. The Qualitative Process Engine, University of Illinois Tech. Report
11 Guyton, A. C. Medical Physiology, Philadelphia, W.B. Saunders, 1986
13 Pool, D. L. Default Reasoning and Diagnosis as Theory Formation, University of Waterloo Tech. Rept. CS-86-08, March 1986
14 Ruten, R. A Logic for Default Reasoning, Artificial Intelligence, April 1980, 14
15 Ruten, R. A Theory of Diagnosis from First Principles, Artificial Intelligence, April 1987, 32(1), 57-96