9-1-1991

Towards Goal-Directed Diagnosis (Preliminary Report)

Ron Rymon
University of Pennsylvania

Bonnie L. Webber
University of Pennsylvania, bonnie@inf.ed.ac.uk

John R. Clarke
Medical College of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-67.

Towards Goal-Directed Diagnosis (Preliminary Report)

Abstract

Recent research has abstracted diagnosis away from the *activity* needed to acquire information and to act on diagnosed disorders. In some problem domains, however, such abstraction is counter-productive and does not reflect real-life practice, which *integrates* diagnostic and therapeutic activity. Trauma management is a case in point. Here, we discuss a formalization of the integrated approach taken in TraumAID, a system we have developed to serve as an artificial aide to residents and physicians dealing with multiple trauma.

Among other things, the active pursuit of information raises the question of what is and what is not worth pursuing. In TraumAID 2.0, we take the view that the process of diagnosis should continue only as long as it is likely to make a difference to future actions. That view is formalized in the *goal-directed* diagnostic paradigm (GDD). Unlike other diagnostic paradigms, goal-directed diagnosis is first and foremost concerned with setting goals based on its conclusions. It regards the traditional construction of an explanation for the faulty behavior as secondary.

In order to explicitly represent goal-directedness, the diagnostic *process* is viewed as search in a space of attitude-beliefs. From this, we derive a high-level algorithm that produces appropriate requests for action *while* searching for an explanation. A complete explanation, however, is not the criterion for terminating action. Such a criterion, we argue, is better treated in terms of goal-means tradeoffs. TraumAID's architecture, in so far as it embodies this goal-directed approach, assigns to a complementary *planner* the resolution of such tradeoffs.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-67.

Towards Goal-Directed Diagnosis
(Preliminary Report)

MS-CIS-91-67
LINC LAB 208

Ron Rymon
(University of Pennsylvania)
Bonnie L. Webber
(University of Pennsylvania)
John R. Clarke
(Medical College of Pennsylvania)

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

September 1991

Towards Goal-Directed Diagnosis (Preliminary Report)

Ron Rymon
University of Pennsylvania
rymon@linc.cis.upenn.edu

Bonnie L. Webber
University of Pennsylvania
bonnie@cis.upenn.edu

John R. Clarke*
Medical College of Pennsylvania
jclarke@grad1.cis.upenn.edu

Abstract

Recent research has abstracted diagnosis away from the *activity* needed to acquire information and to act on diagnosed disorders. In some problem domains, however, such abstraction is counter-productive and does not reflect real-life practice, which *integrates* diagnostic and therapeutic activity. Trauma management is a case in point. Here, we discuss a formalization of the integrated approach taken in TraumAID, a system we have developed to serve as an artificial aide to residents and physicians dealing with multiple trauma.

Among other things, the active pursuit of information raises the question of what is and what is not worth pursuing. In TraumAID 2.0, we take the view that the process of diagnosis should continue only as long as it is likely to make a difference to future actions. That view is formalized in the *goal-directed* diagnostic paradigm (GDD). Unlike other diagnostic paradigms, goal-directed diagnosis is first and foremost concerned with setting goals based on its conclusions. It regards the traditional construction of an explanation for the faulty behavior as secondary.

In order to explicitly represent goal-directedness, the diagnostic *process* is viewed as search in a space of attitude-beliefs. From this, we derive a high-level algorithm that produces appropriate requests for action *while* searching for an explanation. A complete explanation, however, is not the criterion for terminating action. Such a criterion, we argue, is better treated in terms of goal-means tradeoffs. TraumAID's architecture, in so far as it embodies this goal-directed approach, assigns to a complementary *planner* the resolution of such tradeoffs.

1 Introduction

In many domains, it is common to distinguish reasoning and activity concerned with *what* problems need be addressed from that concerned with *how* to address those problems. As

*This work has been supported in part by the Army Research Organization under grant DAAL03-89-C0031PRI.

such, AI subsumes as separate sub-disciplines, *diagnosis* research, which concerns itself with locating the source (or sources) of a system's faulty behavior, and *planning* research, which is concerned with the construction of appropriate plans for addressing given goals. For the most part, research on diagnosis ignores any corrective action that may follow, while planning research ignores the reasoning and activity involved in determining its goals and verifying their achievement.

One of the problems with general theories of diagnosis has been that every too often, an exponential number of hypothetical failures (diagnoses) can "explain" the faulty behavior [Rymon 91]. Thus one faces the problem of which possibility to attend to first. [Poole and Provan 90] were the first to note that the optimality of a diagnosis must depend on *post-diagnosis* goals. To that end, [Provan and Poole 91] advocates the use of utilities in order to choose among different potential diagnoses.

In medicine, it has always been recognized that diagnosis and therapy are strongly tied. Thus, medical Artificial Intelligence systems, as early as MYCIN [Shortliffe 74], have always considered them together[1]. Nevertheless, the relationship between diagnosis and therapy planning was often left informal and implicit.

Only recently has work formalizing diagnosis and repair *together* begun to appear within the diagnosis community. [Rushby and Crow 91] have extended Reiter's consistency-based theory of diagnosis to deal with issues of repair. Within that approach the user can define an acceptable mode of operation, and the theory provides for which forms of repair may entail a situation that is compatible with such a requirement. [Friedrich et al 91] presents a general theory for diagnosis and repair. Within that theory, the user may again define acceptable working conditions for the ailing system. A possible-worlds approach is then taken to formally define those steps that can bring the faulty system to such a condition.

In trauma, as in other domains, management can be roughly decomposed into its diagnostic and therapeutic parts. However, due to the urgent nature of trauma management, it is often impossible to defer treatment until after diagnosis is complete. Thus, not only do diagnosis and therapy have an impact on each other, they must also be temporally interleaved. Another important feature of the diagnostic process in trauma management is that *activity* is often necessary in both its diagnostic and therapeutic parts. Such activity is often costly and may present direct risk to the patient.

Goal-Directed diagnosis begins from the point of view that diagnosis is only worthwhile only to the extent that it can affect decisions concerning actions. As such it differs from other characterizations of diagnosis (e.g. [Reggia et al 85a, Reiter 87]), that seek complete explanations for observed faulty behavior. This view is embodied in TraumAID 2.0 — a system designed to serve as an artificial aide to residents and physicians dealing with multiple penetrating injuries in a trauma domain [Webber et al 91].

TraumAID 2.0 integrates diagnostic reasoning with planning and action. Figure 1 describes its basic cycle of reasoning, planning and action. That cycle begins as initial infor-

[1] We found the following interesting comment in an article by Rennels and Shortliffe: "Although MYCIN is often described as a diagnostic program, its principal motivation was therapy planning". Encyclopedia of Artificial Intelligence, p. 589.

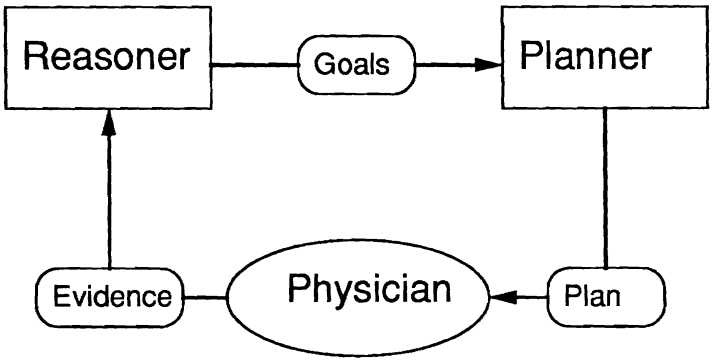

Figure 1: Basic Cycle of Reasoning, Planning and Action

mation (**evidence**) is provided by the **Physician** to the **Reasoner**. From this, the reasoner draws some initial conclusions and suggests preliminary **goals** of action.

Goals are generally categorized as either therapeutic (i.e. that address already diagnosed problems), or diagnostic (i.e. aimed at acquiring information needed to prove or dismiss suspected disorders). In that regard, a goal-directed reasoner is first and foremost dedicated to recommending appropriate treatment goals. The pursuit of diagnostic goals and the explanation of observed faults are strictly secondary, justified only to the extent that they may affect therapeutic decisions.

TraumAID's **Planner** takes those goals and constructs a recommended **plan** of action to address those goals, which it presents to the physician. Results of actions, as well as any new information that becomes available, are reported back to the system to initiate a new cycle of reasoning, planning and action. TraumAID's planner is described in more detail in [Rymon 91a, Webber et al 91, Rymon 90].

An important issue in dynamic diagnostic systems is the timely and orderly acquisition of necessary information. The problem of efficiently acquiring new information along the diagnosis process has been addressed in medical systems [Horvitz et al 84, Shwe et al 89], as well as in domain-independent paradigms [de Kleer and Williams 87, Reggia et al 85b]. Most have taken a probabilistic approach that maximizes the utility associated with the next piece of information. Here, the selection of an appropriate course of action (therapeutic as well as diagnostic) requires resolving goals-means tradeoffs and is therefore cast as a typical planning problem. From a system's point of view, such a characterization results in modularity. More importantly, it presents us with the opportunity to consider the use of a variety of techniques devised for planning problems.

The primary focus of this paper is on the GDD paradigm - a formalized extension of TraumAID's diagnostic reasoning. Section 2 begins by characterizing instances of diagnostic problems, continues with the description of a formal language for specifying such instances, and concludes by defining the diagnostic explanation problem. Section 4 casts diagnostic explanation as a search problem, and derives a meta-algorithm that while searching for an explanation, recommends goals (thus solving the diagnostic problem). Finally, section 5

presents a short example of potential use of the GDD within the TraumAID's architecture.

2 Problem Formulation

Following the intuition of [Reiter 87], a diagnostic explanation problem is defined as one of finding a belief function that is consistent with one's knowledge of the diagnosed system and of the observed (faulty) system behavior. However, for us, this (single) belief constitutes an explanation of that behavior even if it is inconclusive with regard to propositions that are not determined to be of relevance. Next, a language is defined for specifying diagnostic problems in a goal-directed manner, using Prolog-like *evidential* rules to form conclusions from evidence and lower-level conclusions, and *goal-setting* rules to derive diagnostic and therapeutic goals from conclusions or from evidence. Evidential rules are also used to conclude whether recommended goals have been achieved.

2.1 Attitude and Belief

Definition 2.1 *Attitude and Belief Functions*

Given a set of propositions $H \stackrel{\text{def}}{=} \{h_i\}_{i=1}^{n}$,

- *an* **attitude** *function maps H to the set* {R,I} *(relevant and irrelevant).*
- *a* **belief** *function maps H to the set* {T,F,U} *(true, false, or unknown).*
- *an* **attitude-belief** *combines the two and maps H to the set of pairs* $\{\text{R,I}\} \times \{\text{T,F,U}\}$. *Conversely, it can also be viewed as a pair* $\langle A, B \rangle$ *of attitude and belief functions.*

We shall say that an attitude-belief function is **weakly grounded** if its range does not include ⟨R,U⟩. We shall say that it is **strongly grounded** if belief is restricted to {T,F}.

Definition 2.2 *Belief Predicates*

Let h be any proposition, $\langle A, B \rangle$ *any attitude-belief function. Let*

`true(h)` *(or simply h)* $\stackrel{\text{def}}{=}$ $(B(h) = \text{T})$;
`false(h)` $\stackrel{\text{def}}{=}$ $(B(h) = \text{F})$; *and*
`unknown(h)` $\stackrel{\text{def}}{=}$ $(B(h) = \text{U})$.

We also define the following hybrid predicates:

`unless(h)` $\stackrel{\text{def}}{=}$ `false(h)` ∨ `unknown(h)`;
`compatible-with(h)` $\stackrel{\text{def}}{=}$ `true(h)` ∨ `unknown(h)`; and
`known(h)` $\stackrel{\text{def}}{=}$ `true(h)` ∨ `false(h)`.

The predicates `true` and `false` will be called *conclusive* predicates since our inference procedure does not allow negation by failure for antecedents for which we have no confirming nor dismissing information. Similarly, we shall say that our belief is conclusive with regard to a particular proposition h if `known(h)` holds.

Definition 2.3 *Attitude Predicates*

Let h be any proposition, $\langle A, B \rangle$ any attitude-belief function. Let

$\texttt{relevant(h)} \stackrel{\text{def}}{=} (A(h) = \text{R})$*; and*
$\texttt{irrelevant(h)} \stackrel{\text{def}}{=} (A(h) = \text{I})$.

Both `relevant` and `irrelevant` are taken to be conclusive predicates.

2.2 Representing Knowledge

Definition 2.4 *Rules*

A rule ties a conjunction of predicates over propositions to a conclusion or a goal:

1. **Evidential rules** *map evidence (and lower-level conclusions) to conclusions:*

$$Ant_1 \wedge Ant_2 \wedge \ldots \wedge Ant_r \Rightarrow d$$

2. **Goal Setting rules** *map evidence and conclusions to goals:*

$$Ant_1 \wedge Ant_2 \wedge \ldots \wedge Ant_r \rhd g$$

Terms on the left-hand side of a rule will be called **antecedents** and are typically comprised of a predicated proposition (see example 2.5). The right-hand side of the rule is called the **header**, and is a proposition referring to a conclusion or a goal.

We shall say that a rule R **succeeds** if all its antecedents are known to be consistent with their respective predicates. R **fails** only if at least one of its antecedents is known not to hold (i.e. failure does not follow from missing information).

Turning to *semantics*, a goal-setting rule is used to determine one's *attitude* toward the proposition in its header. If the rule succeeds, the proposition is concluded to be relevant; otherwise it is regarded as irrelevant to the problem at hand. Similarly, an evidential rule is associated with one's belief in its header proposition. It differs in that for one to believe that a proposition is false, *all* rules for the particular proposition must fail. Barring that, it will remain unknown.

We allow the same proposition to serve as the header of both goal-setting and evidential rules. In particular, as header to a goal-setting rule, a diagnostic or therapeutic goal means that the rule is used to conclude that the goal is worth adopting. An evidential rule whose header is that same goal is used to conclude whether or not it has been satisfied. Similarly, a goal-setting rule whose header is a clinical condition is used to conclude that it is relevant to investigate that condition. A similarly headed evidential rule is used to conclude whether or not the condition holds. For example, the following evidential rule is used to conclude that a patient's shock is due to abdominal bleeding.

Example 2.5 *Evidential Rule*

```
Shock ∧
false(Single_Wound_to_Upper_Chest) ∧
unless(Pericardial_Tamponade) ∧
unless(Massive_Hemothorax) ∧
unless(Tension_Pneumothorax) ⇒ Shock_of_Abdominal_Origin
```

Next, a goal-setting rule concluding that it is relevant to know whether or not the patient has hematuria.

Example 2.6 *Goal-setting Rule*

```
Wound(Type='Gunshot) ∧
Bullet_in_Abdomen ▷ Hematuria
```

2.3 A Diagnostic Problem

Following Reiter, a diagnostic explanation problem is defined whenever one's beliefs, based on current observations and a knowledge of the underlying system, are inconsistent. In our goal-directed paradigm, on the other hand, a solution to a diagnostic problem is the set of recommendations that is based on such an explanation. Inconsistent beliefs in fact will be "fixed" when appropriate action is taken.

Definition 2.7 *A Diagnostic Problem Instance*

An **instance** *of a diagnostic problem is a 4-tuple* $\langle H, M_0, RB, OBS \rangle$ *such that:*

- $H=\{h_1, h_2, \ldots, h_n\}$ *is a set of propositions.* For consistency with other definitions of diagnostic problems, consider H = DUMUH', where D is a set of disorders, M is a set of manifestations and H' is another set of miscellaneous propositions (such as intermediate conclusions)*;*
- $M_0 \subseteq M$ *is a set of observed manifestations* (i.e. propositions for which we can, initially, assert either `true(m)` or `false(m)`)*;*
- *RB is a set of evidential and goal-setting rules;*
- $OBS : M_0 \rightarrow \{\mathrm{T},\mathrm{F}\}$*, is a partial belief function.*

Definition 2.8 *Consistency*

Given a problem instance $P=\langle H, M_0, RB, OBS \rangle$*, we shall say that an attitude-belief function* $\langle A, B \rangle$ *is* **consistent** *with P, if the following conditions hold:*

1. *B coincides with OBS on* M_0 *(i.e.* $\forall\ m \in M_0\ B(m)=OBS(m)$*).*

2. *for any $h \in H - M_0$ and for any evidential rule $R \in RB$ for which h is a header, whenever R succeeds, $B(h)$=*T.

3. *for any $h \in H - M_0$, if all evidential rules for which h is a header fail, $B(h)$=*F.

4. *for any $h \in H$, $A(h)$=*R *if there is a goal setting rule R in RB, such that h is a header of R and R succeeds; otherwise $A(h)$=*I*:*

The above definition provides a semantic interpretation for rules. Note that, under this interpretation, hybrid predicates (i.e., ones that can function as either conclusions or goals) are just syntactic sugar: any rule that contains such a predicate can easily be transformed into two rules that do not.

Definition 2.9 *Candidate Diagnosis*

Given a problem instance $P=\langle H, M_0, RB, OBS\rangle$, a belief function Δ is a **candidate diagnosis** *for P if there exists an attitude function A such that $\langle A, \Delta\rangle$ is consistent w.r.t. P.*

We shall say that Δ is **weakly complete** if $\langle A, \Delta\rangle$ is weakly grounded (i.e. no relevant propositions are unknown), and **strongly complete** if it is strongly grounded (i.e. all propositions are known).

Definition 2.10 *Refining Beliefs*

Let $\langle A_1, B_1\rangle$ and $\langle A_2, B_2\rangle$ be attitude-belief functions and let h be a proposition. We say that $\langle A_2, B_2\rangle$ is an **immediate refinement w.r.t. h** *of $\langle A_1, B_1\rangle$ (denoted $\langle A_2, B_2\rangle \sqsubset_h \langle A_1, B_1\rangle$) iff*

1. *for all $h' \in H - \{h\}$, $\langle A_2, B_2\rangle$ coincides with $\langle A_1, B_1\rangle$; and*

2. *for h, either of the following holds:*

   *(a) $A_2(h)=A_1(h)$, $B_2(h) \in$ {T,F} and $B_1(h)$=*U*; or*

   *(b) $\langle A_2, B_2\rangle(h)=\langle$*I,U*$\rangle$ and $\langle A_1, B_1\rangle(h)=\langle$*R,U*$\rangle$.*

Figure 2 illustrates this second point schematically, with the direction of immediate refinement depicted as downward. Informally, a concrete (conclusive) belief is more refined than one that is not, and a non-concrete belief is more refined if its inconclusiveness is limited to irrelevant proposition. Note that the $\sqsubset_h$ relation is transitive and anti-symmetric and so can be viewed as a partial order.

Let $\sqsubset \stackrel{\text{def}}{=} \cup_{h \in H} \sqsubset_h$ then we shall say that $\langle A_2, B_2\rangle$ is an **immediate refinement** *of $\langle A_1, B_1\rangle$ (without referring to a particular proposition) iff $\langle A_2, B_2\rangle \sqsubset \langle A_1, B_1\rangle$.*

Let $\sqsubset^*$ denote the transitive closure of $\sqsubset$. We shall say that $\langle A_2, B_2\rangle$ is a **refinement** *of $\langle A_1, B_1\rangle$ iff $\langle A_2, B_2\rangle \sqsubset^* \langle A_1, B_1\rangle$.*

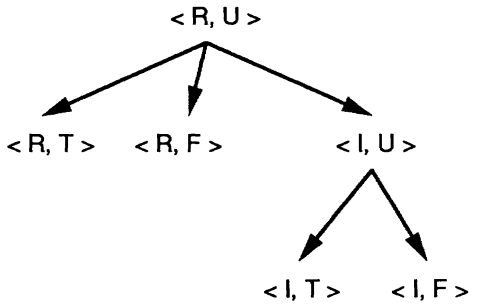

Figure 2: Partial Order Induced by the Refinement Relation

Proposition 2.11 *A strongly complete diagnosis cannot be further refined. A weakly complete diagnosis can only be refined with respect to irrelevant propositions.*

Definition 2.12 *An Explanation to a Diagnostic Problem*

A strong (weak) **explanation** *for a problem instance is a strongly (weakly) complete diagnosis.*

By our GDD premise, active diagnostic refinement shall be restricted to those issues that can affect future decision. The meta-algorithm presented in section 4.2 thus excludes strong explanation as one of its goals.

3 Goal Inhibition

Trauma management presents an on-going interplay between diagnosis and therapy. For example, it is often necessary to delay or even ignore a diagnostic goal in order to perform an immediate therapy. A therapeutic goal may have to give way to other such goal/s that have a broader or more important coverage. Logistic considerations may call for delaying a therapeutic procedure in order to be able to perform a certain, related or not, diagnostic test and so on.

As previously explained, trauma management considers a test *legitimate* only if it can differentiate diagnoses that require different treatment. However, legitimate tests may also have to be deferred until more urgent problems are resolved.

While issues of goal precedence are addressed by the planner, we do not want the diagnostic reasoner to *propose* inappropriate goals. A diagnostic goal is inappropriate if, for example:

1. The condition being diagnosed has already been proved or dismissed by other means;
2. Its treatment, if positively diagnosed, is already covered by that of another, already diagnosed, problem;

3. It is an intermediate diagnosis and all its potential consequences are either proved, falsified, or irrelevant;

4. It cannot be treated for reasons beyond the diagnostic reasoner's control (e.g. due to lack of equipment, or skill, or due to conflicts with other therapy).[2]

Similarly, a therapeutic goal is inappropriate if it was motivated by a partial diagnosis, and a more specific diagnosis was subsequently reached, or if an overriding condition or therapeutic goal was detected.

One way to achieve goal inhibition is to put conditions in relevant goal-setting rules, in the form of antecedent clauses that cause them to fail when their goal is deemed inappropriate. (This was the case in TraumAID 1.0.) For example, a bullet in the abdomen requires a laparotomy. Once the need for a laparotomy is determined though, it is inappropriate to investigate other conditions which would also motivate a laparotomy: the surgeon will repair all injuries when the laparotomy is performed, not just the one that first motivated it. The obvious problem with this solution is that rules become complex and hard to maintain, while also losing the separate function that these particular antecedent clauses are meant to serve (cf. [Clancey 83].)

Our alternative solution to goal inhibition is a general framework for specifying relationships among goals and between goals and conclusions. It is in the form of two hierarchies that can be expanded over a set of GDD rules before they are run, thereby adding relevant inhibition clauses automatically.

In general, there are two classes of goal inhibitions based on goal scaling (or more generally goal hierarchy). First, the pursuit of a goal may be terminated when all higher level goals for which it serves have already been concluded (either true or false). Secondly, one may eliminate a goal if a higher level goal has been concluded as relevant (e.g. all therapeutic and diagnostic goals are overridden when a need for *emergency thoracotomy* is concluded).

Definition 3.1 *Goal Inhibition Hierarchies*

A goal inhibition hierarchy is a partial order on goals. In a conclusion-based inhibition hierarchy, if g and h are goals and g<h, then whenever h is concluded (either true or false), g becomes irrelevant. In a relevance-based inhibition hierarchy, whenever h becomes relevant, g becomes irrelevant.

As noted before, this goal hierarchy can be specified declaratively and then implemented as a macro expansion over the presented specification language. Essentially, for each g, h, such that $g < h$, the macro expansion adds to each goal-setting rule headed by g a clause stating `unknown(h)` for relations from the conclusion-based hierarchy, and `irrelevant(h)` for relations from the relevance hierarchy.

[2]This latter constraint is not within TraumAID's purview. While the choice of procedures recommended by TraumAID's planner is sensitive to currently available resources and physician preferences, physicians will always be informed if a motivated goal cannot be satisfied by any means currently available.

4 The Diagnostic Process

4.1 Diagnosis as Search

Diagnosis can be viewed as a search through a space of states, each corresponding to a particular attitude-belief function, that are linked on the basis of the above refinement relation. Explaining a diagnostic problem is equivalent to a search from an identified initial state – one that best describes the initial knowledge when diagnosis commences – to a goal state – corresponding to an acceptable explanation. Transitions in this space will correspond to change in information, or attitude toward particular pieces of information[3].

Definition 4.1 *States*

Formally, a **state** *is a pair* $\langle H, \Gamma \rangle$*, where* H *is the set of propositions and* Γ *is an attitude-belief function on* H*. However, since for a particular problem* H *is fixed, it allows speaking of states and attitude-beliefs interchangeably.*

States are accessed where there is a change in attitude or belief. In a diagnostic process, a change in an attitude or a belief with regard to a certain proposition will most often be the result of the availability of new information. It is thus necessary to show that these arcs can reflect any possibly required updates in one's attitude-belief.

Definition 4.2 *Updated Attitude-Belief*

Let $\Gamma \stackrel{\text{def}}{=} \langle A, B \rangle$ *be an attitude-belief function. We define* $\Gamma\,|_{B(h)=v}$ *to be an attitude-belief that is the same as* Γ *except that the belief about* h *is updated to* v *(where* $v \in \{\text{T},\text{F},\text{U}\}$*). Similarly, we define* $\Gamma\,|_{A(h)=v}$ *to be the same as* Γ *except that the attitude toward* h *is updated to* v *(where* $v \in \{\text{I},\text{R}\}$*).*

In order to define an accessibility function *Result*, which will account for all possible one-proposition attitude-belief updates, *Up* and *Down* functions are defined that correspond directly to transitions on the refinement hierarchy.

Definition 4.3 *Up and Down operators*

Given an attitude-belief function $\Gamma \stackrel{\text{def}}{=} \langle A, B \rangle$*, a proposition* h*, and a new attitude or belief (for* h*)* v*, we define:*

$$Up(\Gamma,\ h,\ v) = \begin{cases} \Gamma\,|_{B(h)=U} & \text{if B(h)} = \text{T and v} = \text{U} \\ \Gamma\,|_{B(h)=U} & \text{if B(h)} = \text{F and v} = \text{U} \\ \Gamma\,|_{A(h)=R} & \text{if A(h)} = \text{I, B(h)} = \text{U and v} = \text{R} \\ \text{undefined} & \text{otherwise} \end{cases}$$

[3]Information is taken here in a broad sense, ranging from making another observation, concluding the presence of a disease, or reporting the performance of a particular action.

$$Down(\Gamma,\ h,\ v) = \begin{cases} \Gamma\,|_{B(h)=T} & \text{if B(h) = U and v = T} \\ \Gamma\,|_{B(h)=F} & \text{if B(h) = U and v = F} \\ \Gamma\,|_{A(h)=I} & \text{if A(h) = R, B(h) = U and v = I} \\ \text{undefined} & \text{otherwise} \end{cases}$$

Informally, *Up* corresponds to a transition to a less refined state, while *Down* corresponds to a transition to a more refined state.

Definition 4.4 *Immediate Accessibility*

Immediate accessibility *corresponds to a change in the attitude or belief with regard to a particular proposition h. Let* $\Gamma \stackrel{\text{def}}{=} \langle A, B\rangle$ *be a state, h a proposition and v the new attitude or belief for h. We define Result(*Γ*, h, v) as follows (in terms of Up and Down):*

$\Gamma(h)$	v	*Result(*Γ*, h, v)*
$\langle R, U\rangle$	*T, F or I*	*Down(*Γ*, h, v)*
	R or U	Γ *(no change)*
$\langle I, U\rangle$	*T or F*	*Down(*Γ*, h, v)*
	R	*Up(*Γ*, h, R)*
	I or U	Γ *(no change)*
$\langle R, T\rangle$	*F*	*Down(Up(*Γ*, h, U), h, F)*
	U	*Up(*Γ*, h, U)*
	I	*Down(Down(Up(*Γ*, h, U), h, I), h, T)*
	R or T	Γ *(no change)*
$\langle R, F\rangle$	*T*	*Down(Up(*Γ*, h, U), h, T)*
	U	*Up(*Γ*, h, U)*
	I	*Down(Down(Up(*Γ*, h, U), h, I), h, F)*
	R or F	Γ *(no change)*
$\langle I, T\rangle$	*F*	*Down(Up(*Γ*, h, U), h, F)*
	U	*Up(*Γ*, h, U)*
	R	*Down(Up(Up(*Γ*, h, U), h, R), h, T)*
	I or T	Γ *(no change)*
$\langle I, F\rangle$	*T*	*Down(Up(*Γ*, h, U), h, T)*
	U	*Up(*Γ*, h, U)*
	R	*Down(Up(Up(*Γ*, h, U), h, R), h, F)*
	I or F	Γ *(no change)*

Definition 4.5 *Accessibility*

The **accessibility** *function Result* generalizes Result to a sequence of updates. Given an initial state* $\Gamma \stackrel{\text{def}}{=} \langle A, B\rangle$ *and a sequence of updates* $\{h_i, v_i\}_{i=1}^{n}$*, we define*

$$Result^*(\Gamma, \{h_i, v_i\}_{i=1}^{n}) \stackrel{\text{def}}{=} \begin{cases} Result(\Gamma, h_1, v_1) & \text{if n = 1} \\ Result(Result^*(\Gamma, \{h_i, v_i\}_{i=1}^{n-1}), h_n, v_n) & \text{otherwise} \end{cases}$$

4.2 Diagnosis Interleaved With Planning

So far, we have on one hand defined a specification language for diagnostic problems and on the other hand portrayed diagnosis as a search problem. In this section, we will present a meta-level algorithm for diagnosis in an architecture such as TraumAID (see figure 1). This algorithm takes a problem instance specified in the above language and solves it while searching the refinement-based space.

Our meta-algorithm (Algorithm 4.6) emulates transitions in the refinement-based search space. It begins by setting its initial attitude-belief to coincide with the current set of observations. From that point on, transitions will only be made whenever new information becomes available, or when conclusions are made or retracted by the inference engine. To remain consistent with the search space definitions, all transitions are expressed in terms of the accessibility relation *Result*.

Algorithm 4.6 *A Meta-Level Diagnosis Algorithm*

Program Diagnose (H, M_0, RB, OBS).

{* *initialization of beliefs and attitudes* *}
For all $h \in H$, *do* $A(h)$=I, $B(h)$=U.
For all $m \in M_0$, *do* $B(m)=OBS(m)$.

Until (Plan exhausted) and (No more applicable rules)
    {* *make all inferences. try to reach consistency.* *}
    Until no more applicable rules, do
        Let R be an applicable rule,
            $\langle A, B \rangle \leftarrow$ *Fire(R,*$\langle A, B \rangle$*);*
    end-until

    {* *follow recommendations* *}
    $P \leftarrow$ *Plan (*$\langle A, B \rangle$*);* {* *construct a plan to satisfy current goals* *}
    Execute (P) until the first piece of information (h,v) comes in;
    $\langle A, B \rangle \leftarrow$ *Result(*$\langle A, B \rangle$*,h,v);*
end-until.

Inference here takes the form of a closure-like operation in which rules are fired until no more of them are applicable. We next define rule applicability and describe the rule firing procedure (*Fire*) in terms of the *Result* operator. Note that changes in attitude-belief that result from rule firings may result in other rules becoming applicable, which will themselves subsequently be selected and fired.

Definition 4.7 *Rule Applicability*

Let $\langle A, B\rangle$ *be an attitude-belief and R be a rule, R is* **applicable** *if A(header(R))=*R*, B(header(R))=*U*, and all of R's antecedents can be verified (i.e. all conclusive antecedents contain known propositions).*

By its definition, inconsistency is equivalent to the existence of an applicable rule. Applicable rules are fired via the following procedure:

Algorithm 4.8 *Resolving Inconsistency via Rule Firing*

Procedure Fire (R, $\langle A, B\rangle$*).*

Case
1. R is an evidential rule:
if R succeeds then
$\langle A, B\rangle \leftarrow$ *Result(*$\langle A, B\rangle$*,header(R),*T*);*
else {* *Since R is applicable, it must have failed* *}
if all rules for header(R) have failed then
$\langle A, B\rangle \leftarrow$ *Result(*$\langle A, B\rangle$*,header(R),*F*);*
2. R is a goal-setting rule:
if R succeeds then
$\langle A, B\rangle \leftarrow$ *Result(*$\langle A, B\rangle$*,header(R),*R*);*
end-case

Note that the criterion to terminate diagnosis is the absence of goals, not the completeness of the reached diagnosis. Recall that it is the instantiation of appropriate therapeutic goals that is important to a goal-directed diagnostician, not necessarily the completeness of the by-product explanation. Avoiding a completeness-oriented criterion establishes the claim that refinement *must* be motivated.

However, also note that *any* information, whether it has been called for or not, whether it is acquired via diagnostic or via therapeutic activity, will be used by the algorithm to refine its current diagnosis and possibly trigger new goals.

4.3 Complexity

To estimate the overall time required for diagnosis, consider the following factors:

1. Initialization. Since we have n propositions this would not take more than $O(n)$.
2. Rule firing. Assuming monotonic change in information (i.e. that a fact reported as true is not retracted later on), each rule cannot be fired more times than the number of its antecedents. More often than not, a rule will be fired only once throughout a

complete session. However, the use of inconclusive antecedents may require re-firing of a rule up to r times, where r is the rule's arity. It is easy to verify that each call to the *Fire* routine requires no more than a constant time and so the overall runtime required for rule firing is of the order of the size of RB – still linear in the problem size.

3. Finally, there is the time taken for planning. As we all know, planning is computationally costly. Even very costly. However, that cost is inescapable since TraumAID must anyhow plan for therapeutic reasons. [Rymon 90] describes a greedy planning paradigm used by TraumAID's planner.

5 Example

This example is meant to illustrate the diagnostic reasoning process just described and the way it complements the activities recommended by a planner such as TraumAID's. It depicts diagnosis and treatment of *hemothorax* problems – internal bleeding in the chest cavity.

Consider a patient arriving at an emergency room in a stable condition, suffering a gunshot wound to the left chest. A new diagnostic problem is thus instantiated with the following two observations:

1. OBS(`Shock`)=F, since the patient is stable;
2. OBS(`Wound(Location='Chest,Side='Left)`)=T

Let $\Gamma \stackrel{\text{def}}{=} \langle A, B \rangle$ denote the system's current attitude-belief. Initially A(h)=I, B(h)=U, for all propositions h∈H. However, as soon as the observations are reported, B changes its value for *Shock* to F, and for `Wound(Location='Chest,Side='Left))` to T. However, $\langle A, B \rangle$ is now inconsistent due, in part, to the following goal-setting rule:

(1) `Wound(Location='Chest,Side='Left)` $\triangleright$ `Simple_Hemothorax(Side='Left)`

Consistency requires *firing* this rule to set A(`Simple_Hemothorax(Side='Left)`)=R.

At this point, if no other issues arise, the system's attitude-belief is consistent. However, since it is not refined with respect to $\sqsubset_h$, for `h=Simple_Hemothorax(Side='Left)`, $\langle A, B \rangle$ is not a complete diagnosis. Hence, the diagnostic goal of proving or dismissing `h` is set and passed along to the planner. A planner such as TraumAID's current planner would recommend a *Survey Chest X-Ray* as a means of obtaining the desired information.

Suppose the physician orders an X-Ray, reporting signs of hemothorax and a compound fracture to the left ribs. While the latter information had not been solicited, nevertheless the system's attitude-belief toward both propositions will be updated to T. (If X-Ray reports are assumed to be complete, beliefs about all other features of the X-ray will be updated to F.) While each of those updated beliefs may trigger further investigation, for this example we will ignore all but the hemothorax finding. The latter triggers the evidential rule:

(2) X_Ray_Simple_Hemothorax(Side='Left) $\Rightarrow$ Simple_Hemothorax(Side='Left)

Note that the change in belief for Simple_Hemothorax(Side='Left) from U to T may also be interpreted as a success in satisfying the knowledge goal of finding out about it, set by rule 1. Note too that we must distinguish a hemothorax finding from the condition of having a hemothorax, since the condition can be diagnosed in other ways, such as through the presence of decreased breathe sounds. (This latter method would only be used if an X-ray machine was not available.)

(3) False(Radiography_Available) $\wedge$
Decreased_Breath_Sounds(Side='Left)
$\Rightarrow$ Simple_Hemothorax(Side='Left)

The presence of a hemothorax triggers the following goal-setting rule:

(4) Simple_Hemothorax(Side='Left) $\triangleright$ Rx_Simple_Hemothorax(Side='Left)

The attitude toward the therapeutic goal Rx_Simple_Hemothorax(Side='Left) is updated from I to R and is referred to the planner. A planner such as TraumAID's would recommend addressing it through the insertion of a chest tube. Evidence that a chest tube has been inserted leads to a goal becoming relevant of ensuring proper placement of the tube and of checking that it is functioning correctly. When both have been verified, the following rule is evaluated to check that the treatment goal for the simple hemothorax was *actually* satisfied:

(5) False(Chest_Tube_Misplaced(Side='Left)) $\wedge$
Chest_Tube_is_Functioning(Side='Left) $\wedge$
Chest_Tube_is_Draining_Blood(Side='Left)
$\Rightarrow$ Rx_Simple_Hemothorax(Side='Left)

In summary, we have tracked the hemothorax from the initial wound report, through its suspicion as more investigation is recommended, continuing with the acquisition of more evidence that allows for concluding its presence and the need to address it, and finally, making sure that the treatment actually works.

6 Summary

We assume that diagnosis is only worthwhile to the extent that it can affect decisions and so have introduced a goal-directed diagnostic paradigm. In contrast to other diagnostic paradigm, goal-directed diagnosis defines a solution to a diagnostic problem as the set of recommendations implied by a diagnosis rather than as the explanation itself.

We have described a language for specifying diagnostic problems in a goal-directed manner, and then defined the notion of an explanation (not a solution) to such a problem. We

have used this definition to derive a meta-algorithm for goal-directed diagnosis. While seemingly, the algorithm searches for an explanation, its true value lies in the goals it generates to facilitate further diagnosis and on-going repair.

Work in progress:

1. Implementation – within the TraumAID project, we have completed a revision of our existing system [Webber et al 91]. The new system's architecture is identical to the one just described. Its diagnostic component is similar, although not identical, to the one described.

   In the last step of the revision process, funded by AHCPR, we have just finished validating the system against 234 theoretical cases. TraumAID will next be tested against a set of 100 actual cases from MCP's trauma center. Another set of 200 cases will then be evaluated by a panel of national experts.

2. Modeling change in beliefs and goals – Temporal projection is the subject of a whole research within the TraumAID group. In addition to that research, within the GDD paradigm, it appears worthwhile to add a time component to beliefs, augmenting the inference rules appropriately with mappings from the times in which the antecedents hold to the times in which the consequent holds. A mechanism of the type suggested by [Console and Torasso 90] seems a natural start point for that process.

References

[Clancey 83] Clancey, W. J., The Epistemology of a Rule-Based Expert System – A Framework for Explanation. *Artificial Intelligence*, 20, 1983, pp. 215-251.

[Console and Torasso 90] Console, L. and Torasso, P., An Approach to Diagnosis on Causal-Temporal Models. *Proc. AAAI Spring Symposium on AI and Medicine*, March 1990, Stanford, CA, pp. 37-41.

[de Kleer and Williams 87] de Kleer, J. and Williams, B., C., Diagnosing Multiple Faults. *Artificial Intelligence*, 32, 1988, pp. 97-130.

[Friedrich et al 91] Friedrich, G., Gottlob, G., and Nejdl W., Towards a Theory of Repair Process. *Proceedings of Model-Based Diagnosis Workshop*, AAAI-91, Anaheim, CA, 1991.

[Horvitz et al 84] Horvitz, E. J., Heckerman, D. E., Nathwani, B. N. and Fagan, L. M., Diagnostic Strategies in the Hypothesis-Directed PATHFINDER System. *First Conference on AI Applications*, 1984, pp. 630-636.

[Poole and Provan 90] Poole, D. and Provan G. M., What is an Optimal Diagnosis? *Conference on Uncertainty in Artificial Intelligence*, pp. 46-53, 1990.

[Provan and Poole 91] Provan G. M. and Poole, D., The Utility of Consistency-Based Diagnostic Techniques. *Proceedings of KR-91*, Cambridge, MA, 1991, pp. 461-472.

[Reggia et al 85a] Reggia, J. A., Nau, D. S. and Wang, P. Y., A Formal Model of Diagnostic Inference. I. Problem Formulation and Decomposition. *Information Sciences 37*, 1985, pp. 227-256.

[Reggia et al 85b] Reggia, J. A., Nau, D. S., Wang, P. Y. and Peng, Y., A Formal Model of Diagnostic Inference. II. Algorithmic Solution and Application. *Information Sciences 37*, 1985, pp. 257-285.

[Reiter 87] Reiter, R., A Theory of Diagnosis From First Principles. *Artificial Intelligence*, 32, 1987, pp. 57-95.

[Rushby and Crow 91] Rushby, J. and Crow, J., Model-Based Reconfiguration: Toward an Integration with Diagnosis. *Proceedings of AAAI-91*, Anaheim, CA, 1991, pp. 836-841.

[Rymon 90] Rymon, R., Webber, B. L. and Clarke, J. R., Progressive Horizon Planning. *7th Israeli Conference for Artificial Intelligence and Computer Vision.* Ramat Gan, 1990. pp. 99-112.

[Rymon 91] Rymon, R., A Final Determination of the Complexity of Current Formulations of Model-Based Diagnosis (Or Maybe Not Final?). *Proceedings of Model-Based Diagnosis Workshop*, AAAI-91, Anaheim, CA, 1991.

[Rymon 91a] Rymon, R., Ph. D. Proposal (forthcoming). LINC Lab, Computer and Information Science, University of Pennsylvania.

[Shortliffe 74] Shortliffe, E. H., MYCIN: A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection. Ph. D. Thesis, Stanford, CA, 1974.

[Shwe et al 89] Shwe M., Blackford, M., Heckerman, D., Henrion, M., Horvitz, E., Lehman, H. and Cooper, G., A Probabilistic Reformulation of the Quick Medical Reference System. *SCAMC-90*, Symposium on Computer Applications in Medical Care, November 1990, Washington D.C., pp. 790-794.

[Webber et al 91] Webber, B. L., Rymon, R. and Clarke, J. R., Flexible Support for Trauma Management through Goal-directed Reasoning and Planning. To appear in *Artificial Intelligence in Medicine.*