A THEORY OF DIAGNOSIS FOR INCOMPLETE CAUSAL MODELS

Luca Console, Daniele Theseider Dupre', Pietro Torasso

Dipartimento di Informatica - Universita` di Torino
Corso Svizzera 185 - 10149 Torino (Italy)

ABSTRACT

One of the problems of the recent approaches to problem solving based on deep knowledge is the lack of a formal treatment of incomplete knowledge. However, dealing with incomplete models is fundamental to many real-world domains. In this paper we propose a formal theory of causal diagnostic reasoning, dealing with different forms of incompleteness both in the general causal knowledge (missing or abstracted knowledge) and in the data describing a specific case under examination. Different forms of nonmonotonic reasoning (hypothetical and circumscriptive reasoning) are used in order to draw and confirm conclusions from incomplete knowledge. Multiple fault solutions are treated in a natural way and parsimony criteria are used to rank alternative solutions.

1. INTRODUCTION

In recent years many approaches have been proposed to provide problem solvers with some form of "deep knowledge". Most of the efforts towards the logical formalization of diagnostic problem solving have been based on "ontological approaches" [9,24] while in the machine learning community the use of a "domain theory" has gained popularity [16]. One of the critical problems in both cases concerns the completeness of the domain model. However, while some attempts to deal with incomplete (and imperfect) theories are being made in the learning community [23], the completeness of the model is a common assumption in first principles diagnostic systems (which have been successfully applied to the solution of problems such as electronic troubleshooting [8, 13]). This assumption is not adequate for the application of ontological approaches to complex problems (such as medical diagnosis or mechanical troubleshooting) [15] since in these cases a complete model is either not available or intractable. Working at multiple levels of abstraction can reduce the computational complexity of the reasoning process; one cannot assume, however, that the lowest level of description is complete[1]. A further limitation of ontological approaches directly concerns the definition of diagnosis as a set of faulty components to be replaced. This definition can be too restrictive since we may want to distinguish between different faults for the same component, and, on the other hand, in many cases a faulty behavior is not due to a small set of responsible faulty components (consider, for example, human "systemic diseases").

Since the early 80's, another major family of approaches to diagnosis has been widely adopted, especially in medical domains: the so called "causal modeling approaches"[2] (after Patil's seminal work [18] and successive proposals as [11,26]). Although widely applied to many real world problems, such approaches seem to be less constrained and well-defined than ontological ones; therefore their formal definition has not yet received much attention. Moreover, some form of incompleteness must be accommodated in causal models (as well as in ontological ones). Although this need has been recognized [17] and has led some researchers to provide probabilistic foundations to causal diagnostic reasoning [7, 19], a lot of work still has to be done on the logical treatment of incompleteness in causal models.

The aim of this paper is to describe a causal modeling formalism which on the one hand accommodates some forms of incompleteness and on the other hand has a precise logical formalization. In particular, the main goals and issues we want to address can be synthesized by the following items:

- introducing a formalism to represent *incomplete causal knowledge*;
- giving a *logical foundation* of causal reasoning on incomplete knowledge and introducing precise definitions of the concepts of "diagnostic problem" and of "solution to a diagnostic problem";
- providing a formal treatment of *multiple-fault diagnosis* within causal models;
- providing some criteria to select the "best" solution to a

The research described in this paper has been partially supported by grants from MPI 40% (Automated Reasoning Techniques for Intelligent Systems) and CNR of Italy.

[1] "Even the lowest level of explanation in a domain theory may incorporate abstractions [...] Abstractions at the lowest level of a domain theory imply missing knowledge.." ([10] page 543).

[2] The term "causal" has not been used here to mean that these approaches are the only ones based on causal knowledge. Many of the ontological approaches are in fact based on the representation of causal

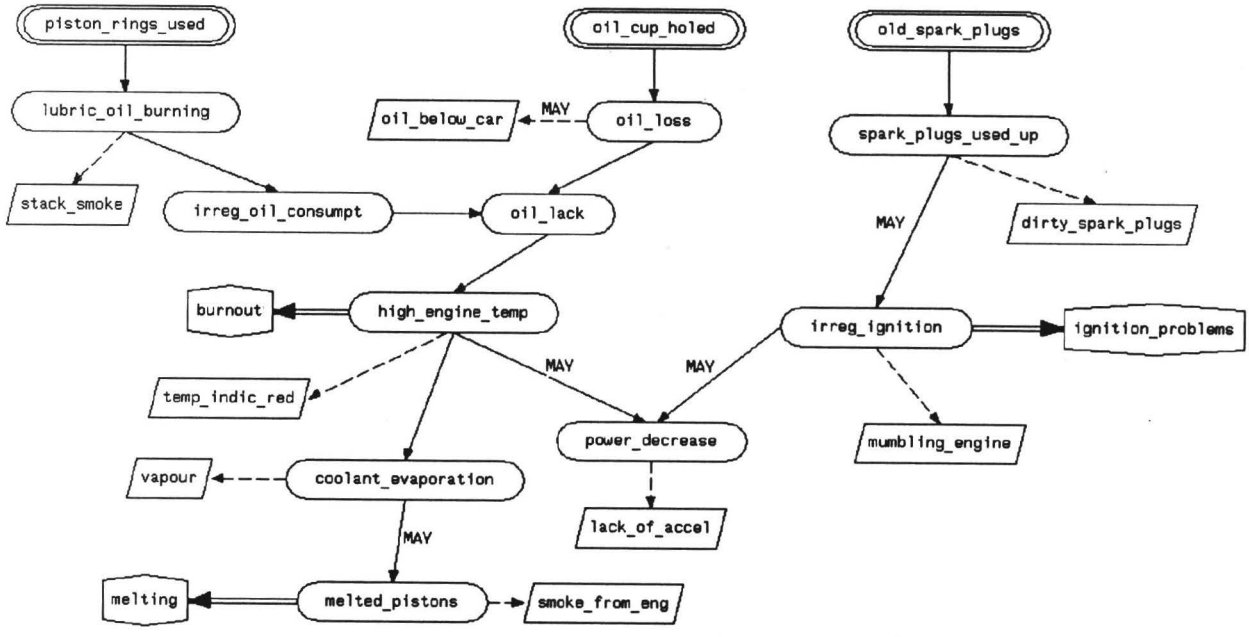

Figure 1 - An example of a causal model in a mechanical domain.

problem.

In our logical approach diagnosis is seen as an *abductive* process starting from the logical correspondent of a causal model. In order to draw conclusions from incomplete models some forms of *nonmonotonic reasoning* are adopted and, as a consequence, the solutions we obtain are defeasible. In particular, instead of using one of the nonmonotonic logics proposed in the literature, we adopt first-order logic supplemented with specific forms of *hypothetical* and *circumscriptive reasoning* [2]. This is in the same line of Poole's first-order approach to default reasoning and diagnosis [21, 22].

This paper is organized as follows: in section 2 we outline an abstract formalism to represent causal knowledge; in section 3 we define a formal theory of causal diagnostic problem solving; in section 4 we present an example of diagnosis.

2. A FORMALISM TO REPRESENT DEEP CAUSAL KNOWLEDGE

In this section we shall briefly introduce a formalism to represent causal knowledge and its logical formalization (the causal modeling formalism we shall introduce is a rather general one, although strongly influenced by our past experiences in the design of the CHECK system [3, 27]). Causal networks are a general formalism to represent causal knowledge, i.e. to describe the set of behaviors of a physical (physiological) system. In the following we shall assume, in particular, that causal networks are used to model the faulty behavior of a system (which seems to be the most interesting case in diagnostic applications).

Consider the simple example in figure 1. At least four types of nodes have to be used in a causal modeling formalism: STATE nodes (elliptic boxes in figure 1), representing partial states of the modeled system; FINDING nodes (rhomboidal boxes), representing observable manifestations of internal states (which cannot be observed); HYPOTHESIS nodes (hexagonal boxes), representing diagnostic hypotheses; INITIAL_CAUSE nodes (double-lined elliptic boxes), representing initial perturbations (i.e. the mechanisms starting the processes which may lead the system to a faulty behavior). In the following we shall assume that INITIAL_CAUSEs are not observable (which is the most general and interesting case). This corresponds to a first form of incompleteness in the model: such nodes are abstractions of the actual perturbation processes, and the model gives no *direct* way to establish their presence.

The nodes in a network can be connected by means of different types of arcs (relationships): each CAUSAL arc (continuous lines in figure 1) connects one or a set (conjunction) of STATE and INITIAL_CAUSE nodes to a STATE node and represents a cause-effect relationship; each HAM arc (dashed lines) connects a STATE node to a FINDING node (HAM - that is Has As a Manifestation - arcs connect states to their external manifestations); each DEFINED_AS arc (double lines) connects a STATE (or a conjunction of STATE nodes) to a HYPOTHESIS node (to represent that the hypothesis is defined as the presence of the state, therefore hypotheses are defined concepts).

In case of different arcs entering a node N, such arcs are implicitly ORed (see, for example, the CAUSAL arcs entering the STATE "oil_lack" in figure 1). Nodes can have attributes (i.e. variables giving a better characterization of the entity represented by the node itself): a function, describing how the values of the attributes of the caused state (manifestation) depend on those of the causing ones, is associated with each CAUSAL and HAM arc. Besides the incompleteness associated with INITIAL_CAUSE nodes, some other form of

knowledge (together with other forms of knowledge).

incompleteness must be accommodated by causal modeling formalisms. Causal models of complex physical (physiological) systems may be incomplete at least for the following reasons (see [2,4] for a discussion):

- Choosing the level of detail of a model is quite arbitrary (or, at least, domain and task dependent). However, in order to constrain the complexity of a model, some bounds have to be imposed to the level of detail at which knowledge is represented [18,23]. Therefore, processes and conditions at a lower level than the chosen one have to be abstracted.
- There exist (especially in physiological systems) processes and conditions which are not perfectly known.

In order to deal with such forms of incompleteness we have introduced the possibility of labeling each CAUSAL or HAM relationship in the causal model as a MAY one. A MAY label denotes that the model of the relationship is incomplete, i.e. that some condition or process has been abstracted. A relationship not explicitly labeled as MAY is assumed to be fully specified.

Let us discuss now how a logical semantic can be associated with the formalism introduced in the previous paragraphs. All the entities and relationships in a causal network can be modeled in a logical framework; the main ideas of such a formalization can be summarized as follows:

- Each STATE, FINDING and INITIAL_CAUSE can be characterized as an atomic formula. More specifically, different sets of predicate symbols are used for each type of node: "state symbols" for STATE nodes, "manifestation symbols" for FINDING nodes and "initial cause symbols" for INITIAL_CAUSE nodes[3]. The arity of a predicate symbol equals the number of attributes of the node it represents. For example, a STATE S with two attributes is represented by the formula S(x,y) with $S \in$ "state symbols".
- CAUSAL and HAM relationships are modeled by implication formulae. *Functions* are used to express how the values of the attributes of the caused state depend on those of the causing ones. For example:
  $$\text{lubric_oil_burning}(x) \rightarrow \text{stack_smoke}(f_1(x))$$
  is part of the logical model of the network in figure 1. The state "lubric_oil_burning" has the attribute "quantity" and the finding "stack_smoke" has the attribute "color"; the function is defined as follows:
  $$f_1(\text{low})=\text{grey and } f_1(\text{high})=\text{black}$$
- MAY relationships are modeled through the introduction of an "abstracted condition symbol"[4]. Each "abstracted condition" atom models the actual condition (process) abstracted (ignored) in the model. As an example from figure 1, we have:
  $$\text{high_engine_temp}(x) \wedge \alpha \rightarrow \text{power_decrease}(f_2(x))$$

Since INITIAL_CAUSEs and MAY relationships represent two forms of incompleteness that will be dealt with in a similar way, we introduce the following definition:

Definition 2.1. The set of **assumption symbols** is the union of the sets of abstracted condition and initial cause symbols, that is:

$$\{\text{assumption symbols}\} = \{\text{abstracted condition symbols}\} \cup \{\text{initial cause symbols}\}$$

A form of "hypothetical reasoning" will be defined in the following sections to deal with such assumptions in the diagnostic reasoning process.

Notice that no logical correspondent is defined for DEFINED_AS relationships. Such relationships are, in fact, at a different knowledge level than the other ones (hypotheses are defined concepts), so it is not necessary to model them as logical formulae. However, since each diagnostic hypothesis is defined as the presence of a conjunction of states (i.e. each HYPOTHESIS node is connected to a set of STATE nodes through a DEFINED_AS arc), we introduce the following definition:

Definition 2.2. Given a diagnostic hypothesis H (i.e. a HYPOTHESIS node H), we denote with **def(H)** the formula

$$\exists\, \overline{x}_1\, S_1(\overline{x}_1) \wedge \ldots \wedge \exists\, \overline{x}_k\, S_k(\overline{x}_k)$$

in case a DEFINED_AS arc connects $S_1,\ldots,S_k$ to H.

In the following, for the sake of brevity, we shall limit ourselves to consider "singly connected" causal networks, i.e. those networks in which, given two nodes N_1 and N_2, there exists at most one path connecting N_1 and N_2. A discussion about more general causal networks can be found in [5].

3. DIAGNOSTIC PROBLEMS AND THEIR SOLUTIONS

In this section we introduce the concepts of diagnostic problem and of solution to a diagnostic problem, defining in which way the solutions are affected by the form of hypothetical reasoning that is used to find them.

3.1. Basic Definitions

A causal network represents the general knowledge about some domain. We have a diagnostic problem when we add to this knowledge the set of findings that have been observed in the case under examination.

Definition 3.1. A **diagnostic problem** (d.p. in the following) P is a triple <NET,HYP,Ψ>, where:

- NET is the set (conjunction) of logical formulae modeling a causal network NETWORK;
- HYP={<H,def(H)> | H is a HYPOTHESIS node in the causal network NETWORK};
- Ψ is a nonempty set of ground manifestation atoms.

Since manifestation predicates represent observable conditions, we assume that the manifestations which have not been declared as observed are absent: therefore in the following we use the set:

$$\overline{\Psi} = \{ \neg m \,/\, m \text{ is a ground manifestation atom, } m \notin \Psi\}$$

[3] In definitions and in symbolic examples "state symbols" will be denoted by upper case roman letters, "manifestation symbols" by lower case roman letters and "initial cause symbols" by the last letters of the greek alphabet.

[4] "abstracted condition" symbols will be denoted by the first letters of the greek alphabet.

This is the first non-monotonic rule we introduce in the diagnostic process in order to draw conclusions from incomplete knowledge: this is a case of (possibly) incomplete factual knowledge, i.e. we only have a partial description of the case under examination. In a real application such a rule must be made explicit, even if it does not preclude that data are acquired incrementally by the system's control strategy (that is, the systems asks for further data when necessary). However, in the following the whole diagnostic process is described in a declarative way, abstracting from the definition of a control strategy. It is worth noting that, as an alternative, we can give the possibility to leave some findings as "unknown"; we shall briefly return to this point at the end of this subsection.

The reasoning process involves the concept of *hypothetical world*, in the (syntactical) sense of a set of formulae including assumptions. Such a world may be regarded as a tentative reconstruction of the causal evolution that has led to the observed situation. The notion of hypothetical world to be used in solving a diagnostic problem is defined recursively. The basis is the case of a world containing initial causes; the recursive case involves abstracted conditions on causal relationships.

Definition 3.2. Given a d.p. P=<NET,HYP,Ψ>, and given a world W, W is a **world** for P iff

$W = NET \cup \{\sigma_1, \ldots, \sigma_k\}$ where σ_i, for i=1,..,k, is a ground initial cause atom (in this case we say that W is an **initial world** for P)

or

$W = W' \cup \{\alpha\}$ where
- W' is a world for P;
- $X \wedge \alpha \rightarrow Y$ is an instance of a formula in NET, where α is an abstracted condition symbol;
- $W' \vdash X$.

In the latter case of the definition, the world $W' \cup \{\alpha\}$ has to be considered in order to examine the consequences of assuming that the condition α is actually true in the case under examination; that is, assuming that the corresponding "MAY" relationship holds in the case under examination.

To test a world against the observed manifestations we introduce the following definitions.

Definition 3.3. Given a d.p. P=<NET,HYP,Ψ>, a world W for P is **inconsistent** iff $W \vdash m$ for some ground manifestation atom m such that $m \notin \Psi$.

In other words, W is inconsistent for P iff $W \cup \overline{\Psi}$ is inconsistent in the logical sense. This has similarities with consistency-based approaches to nonmonotonic reasoning [25], with the difference that in our diagnostic system, which is not a general-purpose approach to nonmonotonic reasoning, the consistency check can be limited to a small class of formulae.

Definition 3.4. Given a world W and a set Ψ of ground manifestation atoms, W **covers** Ψ iff $\forall m \in \Psi,\ W \vdash m$.

In order to give a solution to a diagnostic problem P=<NET,HYP,Ψ>, a world W for P has to explain the observed findings in Ψ and must not predict the presence of findings that have not been observed; this is summarized by the following definition.

Definition 3.5. Given a d.p. P=<NET,HYP,Ψ>, a world W for P is a **final world** iff W covers Ψ and is not inconsistent, that is

$\Psi = \{m \,/\, m \text{ ground manifestation atom}, W \vdash m\}$.

The definition of solution to a diagnostic problem is given as follows, in terms of diagnostic hypotheses whose definition is satisfied in a final world, i.e. in a world representing a causal evolution that is consistent with the observed situation.

Definition 3.6. Given a d.p. P=<NET,HYP,Ψ> and a final world W for P, the set

$\text{diagnosis}(W) = \{ H \,/\, \langle H,\text{def}(H)\rangle \in HYP,\ W \vdash \text{def}(H) \}$

is a **solution** to P. We say that W is a **causal explanation** of the solution.

It is worth noting that this kind of explanation of a solution may be very important for the user, for example to remove the fault in the system under examination.

The definitions in this section are sufficient to characterize diagnosis as a nonmonotonic process. Consider, for example, the network in figure 2. In such a case we have that $\Theta_1=\{H_1\}$ is the solution to the problem characterized by $\Psi_1=\{m_1\}$ and $\Theta_2=\{H_2\}$ is the solution to the problem characterized by $\Psi_2=\{m_1,m_2\}$.

The previous definitions can be extended to the case of "unknown" findings. In such a case the set of ground manifestation atoms is partitioned into the three classes of "observed", "unknown" and "absent" ones. A world W is inconsistent if and only if an "absent" finding is deduced in W, while final worlds must cover only observed findings and may predict the presence of some unknown finding (more details are given in [5]).

3.2. Confirmation of assumptions and solutions

The form of hypothetical reasoning described in the previous definitions is the first step to solve the problem of dealing with incomplete knowledge. It is not enough because, in many cases of interest in real-world applications, there may be more than one final world and then more than one solution to a diagnostic problem, according to the definitions given above. A solution is *one* of the

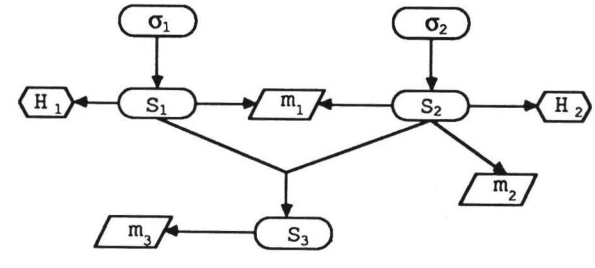

Figure 2 - A simple causal network exemplifying the nonmonotonicity of the diagnostic process.

possible explanations of the given data. The problem is that we have no direct way to establish the truth of the conditions that have been assumed in order to explain the observed findings: we need some stronger form of abductive reasoning to give an *indirect confirmation* to the assumptions, or, at least, to give some preference criteria between the different solutions to a diagnostic problem.

In particular, we introduce a *non monotonic* criterion for the confirmation of assumptions. The conclusions obtained in this way are strongly dependent on the hypothesis that the causal network contains all the relevant knowledge, that is all the possible causes for a given state and all the states that (may) have a given finding as a manifestation. Therefore, such conclusions are defeasible, as it is typical in nonmonotonic reasoning. The idea is to confirm those assumptions that are necessary to explain the presence of the observed findings.

Definition 3.7. Given a d.p. P, an assumption α is **confirmed** iff for every final world W for P, $\alpha \in W$.

The extension of the definition to give confirmation to a world and to a solution is straightforward.

Definition 3.8. A final world W for a d.p. P is **confirmed** iff every assumption $\alpha \in W$ is confirmed.

In other words, a final world W is confirmed if its assumption set Assumptions(W) is a subset of the assumption sets of any (other) final world.

Definition 3.9. A solution Θ to a diagnostic problem P is **strongly confirmed** iff Θ = diagnosis(W) for some confirmed final world W.

This is the most natural case for the confirmation of a solution: it allows us to confirm the explanation of a solution, i.e. to confirm an instance of a subnetwork representing the causal evolution that has led to the observed situation. There are cases where this cannot be done, but it is still possible to give some kind of "weak" confirmation to a solution, directly confirming diagnostic hypotheses without dealing with the (more primitive) level of the assumptions.

Definition 3.10. Given a d.p. P=<NET,HYP,Ψ>, a diagnostic hypothesis H occurring in HYP is **confirmed** iff $H \in \Theta$ for every solution Θ to P.

Definition 3.11. A solution Θ to a diagnostic problem P is **weakly confirmed** iff every $H \in \Theta$ is confirmed (and Θ is not strongly confirmed).

The following property shows that the concept of confirmed solution is well-defined.

Property 1. A diagnostic problem has at most one (either strongly or weakly) confirmed solution.[5]

3.3. Object level characterization of confirmation

We have seen in the previous section that the confirmation of assumptions is based on the non-monotonic hypothesis that all relevant knowledge is modeled. If this is the case, when a state (or manifestation) is present, one of its possible causes must be present. This can be formalized using Mc Carthy's *circumscription principle* [14], giving in such a way an object level characterization of the confirmation criterion. In particular, we use *parallel circumscription* in *ordered formulae* which has been proven equivalent to a first-order formula ([12], theorem 6.11, p.148). Notice that in our case the same results can be obtained using *predicate completion* [1].

Considering the topological order defined in a singly connected causal network, it is easy to prove that the logical model of such a causal network is an ordered formula with respect to the set of state and manifestation predicates.

We give, therefore, an alternative definition for the confirmation of assumptions using the parallel circumscription of the set of state and manifestation predicates in the formula NET - that is, the (first-order) formula NET_{circ} defined as follows, (where STATE and FINDING are the classes of state and manifestation symbols of the alphabet):

$$NET_{circ} = CIRC[NET; STATE \cup FINDING]$$

This formula gives a formalization to the intuition described above and can be used to give the following definition.

Definition 3.12. Given a d.p. P=<NET,HYP,Ψ>, an assumption α is **circumscription-confirmed** iff

$$NET_{circ} \cup \Psi \cup \overline{\Psi} \vdash \alpha$$

The two different criteria can be proven to be equivalent.

Theorem 2. Given a d.p. P=<NET,HYP,Ψ>, an assumption α is confirmed iff it is circumscription-confirmed.

The importance of this result stems from the different nature of definitions 3.7 and 3.12. The latter is an object level characterization, while the former is a meta level one and can be easily implemented, since we are generally interested in finding all the final worlds for a problem (or at least the non redundant ones - see next section). The object-level approach can be generalized to characterize the entire set of solutions to a diagnostic problem [6].

3.4. Parsimonious solutions

The confirmation criteria defined in the previous section may sort out one of the solutions to a diagnostic problem; when there is no confirmed solution, we may still be interested in ranking the solutions, trying to find the "best" one or at least a selected subset of equally good or incomparable ones.

This problem has been analyzed by some researchers over the last few years. In the "Parsimonious Covering Theory" [20] the notions of *minimality* and *minimum cardinality* have been used to define some parsimony criteria. Similarly de Kleer [9] and Reiter [24] proposed parsi-

[5] The proof of this property and of the following ones can be found in [5].

mony criteria based on minimality considerations[6].

The basic justification supporting the "minimal cardinality" parsimony criterion is the assumption that faults are independent. Such an assumption can be questioned in many real world domains. In many cases, in fact, two faults (diseases or diagnostic hypotheses) can be causally correlated (e.g. "burnout" and "melting" in figure 1). On the other hand, the "minimality" (irredundancy) parsimony criterion seems to be much more widely applicable since it does not require any a-priori assumption on the nature of the diagnostic hypotheses being considered. Of course this criterion is weaker, since it is based on a partial order, where we may have different (and incomparable) minimal elements (while in the other case two different solutions of minimum cardinality can be considered equally good).

In our case the two criteria can be applied to the solutions as sets of diagnostic hypotheses, and, more interestingly, to the set of assumptions that characterize the final worlds or, equivalently, to the final worlds themselves. We have already remarked that assumptions are a more primitive notion; moreover, if a solution has more than one explanation, we can select the best one by considering their assumption set.

Definition 4.1. A solution Θ to a diagnostic problem P is
- SA-parsimonious iff it has an explanation which has a $\subseteq$-minimal assumption set within the final worlds for P;
- SH-parsimonious iff it is a $\subseteq$-minimal element of the set of solutions to P;
- CA-parsimonious iff it has an explanation with minimum number of assumptions within the final worlds for P;
- CH-parsimonious iff it has minimum cardinality within the set of solutions to P.

The various definitions are not independent from each other and from the concept of confirmed solution; these relationships are summarized by the following properties.

Property 3. Given a solution Θ to a diagnostic problem,
(a) Θ is SA-parsimonious iff it is SH-parsimonious;
(b) if Θ is CA-parsimonious, then it is SA-parsimonious;
(c) if Θ is CH-parsimonious, then it is SH-parsimonious.

Property 4.
(a) A strongly confirmed solution to a diagnostic problem is CA-parsimonious and CH-parsimonious (and then SH-parsimonious and SA-parsimonious).
(b) A weakly confirmed solution to a diagnostic problem is CH-parsimonious (and then SH-parsimonious and SA-parsimonious).

It is worth noting that there is no general relation between CA-parsimony and CH-parsimony or weak confirmation. Consider the example in figure 3. For the problem characterized by $\Psi_1=\{m_1\}$ the solution $\Theta_1=\{H_1\}$ obtained under the assumption set $\{\sigma_1,\sigma_2\}$ is CH-parsimonious and weakly confirmed; the solution $\Theta_2=\{H_1,H_2\}$ obtained under the assumption set $\{\tau_1\}$ is CA-parsimonious.

[6] The idea in Reiter's and de Kleer's approaches as well as in our one is that parsimony can be also used to gain in efficiency, avoiding the construction of redundant solutions.

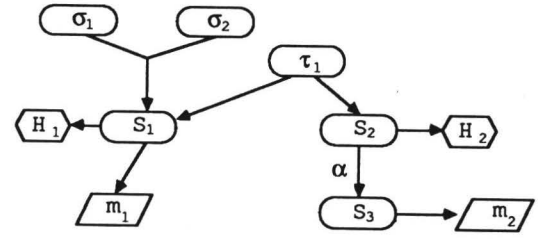

Figure 3 - A simple causal network exemplifying the different concepts of parsimony.

4. AN EXAMPLE

As a simple example in a real-world domain, consider the following diagnostic problem on the causal model in figure 1, where we assume, for the sake of simplicity, that nodes have no attributes:
P_1=<NET_1,HYP_1,Ψ_1>, where:
- NET_1 is the logical model of the network
- HYP_1={<burnout,high_engine_temp>,
  <melting,melted_pistons>,
  <ignition_problems,irreg_ignition>}
- Ψ_1 = {dirty_spark_plugs, lack_of_accel, temp_indic_red, vapour}

The only final world is that containing the assumptions {oil_cup_holed, old_spark_plugs, α}, where α is the assumption associated with the CAUSAL arc between the states "high_engine_temp" and "power_decrease". Therefore the only solution to P is {burnout}. Notice that the presence of "old_spark_plugs" is necessary to explain "dirty_spark_plugs", but the related diagnostic hypothesis "ignition_problems" is not part of any solution, due to the absence of "mumbling_engine". In this case the "MAY" relationship between "spark_plugs_used_up" and "irreg_ignition" does not hold. On the other hand, the initial assumption "oil_cup_holed" is necessary to explain "temp_indic_red" and "vapour"; in fact "stack_smoke" is absent and thus the assumption "piston_rings_used" is rejected while the absence of "oil_below_car" is not relevant because it is not a necessary consequence of "oil_cup_holed".

5. CONCLUSIONS

In this paper we have presented a formal theory of causal diagnostic reasoning, paying particular attention to the problem of dealing with incomplete knowledge. Different forms of incompleteness have been discussed, and a non-monotonic reasoning process for drawing (defeasible) conclusions in the presence of incompleteness has been presented. In summary, such forms of incompleteness can be divided into two main classes:
- *Abstracted knowledge:* initial causes and abstracted

conditions fall within such a class. A hypothetical reasoning scheme has been designed in order to deal with abstracted knowledge.

- *Missing knowledge:* in order to draw conclusions from incomplete models, different forms of circumscriptive reasoning have been used on the causal model itself and on the findings observed in a specific case under examination.

The diagnostic process described in this paper has been implemented in Prolog, experimenting different control strategies in order to avoid inefficiencies. In particular, compiled heuristic rules are used to focus the reasoning process (i.e. to limit the number of initial worlds to be considered) [5].

REFERENCES

[1] K. Clark, "Negation as failure," pp. 293-322 in *Logic and Data Bases*, ed. H. Gallaire, J. Minker,Plenum Press, New York (1978).

[2] L. Console and P. Torasso, "Hypothetical Reasoning in Causal Models," *to appear in: International Journal of Intelligent Systems*, (1989).

[3] L. Console and P. Torasso, "Heuristic and Causal Reasoning in Medical Diagnosis," pp. 16-17 in *Proc of the AAAI Symposium on Artificial Intelligence in Medicine*, Stanford (March 1988).

[4] L. Console and P. Torasso, "A Logical Approach to Deal with Incomplete Causal Models in Diagnostic Problem Solving," pp. 255-264 in *Lecture Notes in Computer Science 313*, Springer Verlag (1988).

[5] L. Console, D. Theseider Dupre´, and P. Torasso, "Hypothetical Reasoning in Diagnostic Problem Solving," working paper (1989).

[6] L. Console, D. Theseider Dupre´, and P. Torasso, "Abductive Reasoning through Direct Deduction from Completed Domain Models," submitted for publication (1989).

[7] G. Cooper, "NESTOR: A computer-based medical diagnostic aid that integrates causal and probabilistic knowledge," Techn. Rept. 84-48, Dept. of Computer Science, Stanford University (1984).

[8] R. Davis, "Diagnostic reasoning based on structure and behavior," *Artificial Intelligence* **24** pp. 347-410 (1984).

[9] J. de Kleer and B.C. Williams, "Diagnosing Multiple Faults," *Artificial Intelligence* **32** pp. 97-130 (1987).

[10] R. Doyle, "Constructing and Refining Causal Explanations from an Inconsistent Domain Theory," pp. 538-544 in *Proc. AAAI 86*, Philadelphia (1986).

[11] P. Fink, J. Lusth, and J. Duran, "A General Expert System design for Diagnostic problem Solving," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-7** pp. 553-560 (1985).

[12] M.R. Genesereth and N.J. Nilsson, *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann (1987).

[13] M.R. Genesereth, "The use of design descriptions in automated diagnosis," *Artificial Intelligence* **24** pp. 411-436 (1984).

[14] J. McCarthy, "Circumscription: a form of non-monotonic reasoning," *Artificial Intelligence*, **13** pp. 27-39 (1980).

[15] D. McDermott, "Logic, Problem Solving and Deduction," pp. 187-229 in *Annual Review of Computer Science, Vol.2*, (1987).

[16] T. Mitchell, R. Keller, and S. Kedar-Cabelli, "Explanation-Based Generalization: A Unifying View," *Machine Learning* **1** pp. 47-80 (1986).

[17] D. Partridge, "The Scope and Limitation of First Generation Expert Systems," *Future Generation Computing Systems* **3**(1) pp. 1-10 (1987).

[18] R. Patil, "Causal representation of patient illness for electrolyte and acid-base diagnosis," MIT/LCS/TR-267, Cambridge (1981).

[19] J. Pearl, "Distributed Revision of Composite Beliefs," *Artificial Intelligence* **33** pp. 173-215 (1987).

[20] Y. Peng and J. Reggia, "A Probabilistic Causal Model for Diagnostic Problem Solving - Part I: Integrating Symbolic Causal Inference with Numeric Probabilistic Inference," *IEEE Trans. on Systems, Man and Cybernetics* **SMC-17** pp. 146-162 (1987).

[21] D. Poole, "A Logical Framework for Default Reasoning," *Artificial Intelligence* **36** pp. 27-47 (1988).

[22] D. Poole, "Representing Knowledge for Logic-based Diagnosis," pp. 1282-1290 in *Proc. of the Int. Conf. on Fifth Generation Computer Systems*, Tokyo (1988).

[23] S. Rajamoney and G. DeJong, "The classification, detection and handling of imperfect theory problems," pp. 205-207 in *Proc. IJCAI 87*, Milano (1987).

[24] R. Reiter, "A Theory of Diagnosis from First Principles," *Artificial Intelligence* **32** pp. 57-96 (1987).

[25] R. Reiter, "Nonmonotonic reasoning," pp. 147-186 in *Annual Review of Computer Science, Vol.2*, (1987).

[26] L. Steels, "Second Generation Expert Systems," *Future Generation Computing Systems* **1** pp. 213-221 (1985).

[27] P. Torasso and L. Console, *Diagnostic Problem Solving: Combining Heuristic, Approximate and Causal Reasoning*, Van Nostrand Reinhold (1989).