Defining and exploring properties in diagnostic systems

Nuno Belard [1] [2] [3], **Yannick Pencolé** [2] [3] **and Michel Combacau** [2] [3]

[1] *Airbus France; 316 route de Bayonne; 31060 Toulouse, France*
[2] *LAAS-CNRS; 7 Avenue du Colonel Roche; F-31077 Toulouse, France*
[3] *Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; 7 Avenue du Colonel Roche, F-31077 Toulouse, France*
nuno.belard@airbus.com , ypencole@laas.fr , combacau@laas.fr

ABSTRACT

Every model-based diagnostic approach relies on a representation of a real-world system, in this paper called believed system. The believed system is used along with the observations about the real-world system to generate a diagnostic problem to be solved. In this paper it is firstly argued that believed systems can differ from real-world systems in many different manners. As so, properties of believed systems, diagnostic problems and diagnostic results are introduced. Then, a series of relations between these properties are proved. The importance of such relations, sometimes seen as intuitive, is that they are necessary to formally prove the accordance between the real-world system and the believed system; to formally prove that a believed system and a diagnostic problem will produce high-quality diagnostic results; or even to ease diagnostic algorithms, since for systems and problems with certain properties, different model-based diagnostic approaches produce the same diagnostic results. In order to introduce the referred properties and reasoning about them a framework of diagnosis based on the difference between the believed and the real systems is proposed.

1 INTRODUCTION

Diagnostic reasoning aims at determining the normal and faulty components of a system under study, typically for repairing the faulty ones, based on a formal representation of such system and on a series of observations about it.

Since this description of diagnostic reasoning is beyond doubt on the line of thought of, for instance, Reiter (Reiter, 1987) and de Kleer and Williams in (de Kleer and Williams, 1987), we start by noticing that these approaches clearly distinguish the "real world setting of interest" (or "artifact" or "physical system") from the description of such system. Let us, in this paper, call the former *real system* and the latter *believed system* [1]. As described, apart from the real and the believed system, diagnostic reasoning also involves *observations* about the real system. So how does a diagnostic reasoner work?

A diagnostic reasoner operates by by using the observations (about the real system) and treat them as being observations about the believed system. Then, this agent of diagnosis computes the possible *health states* of the believed system, that is, those possible assignments of abnormal or normal states to each component of the believed system that, typically, are either consistent with all the observations or, in a much stronger manner, actually imply some of them. Finally, under the [almost always implicit] assumption that the believed system is a "good" representation of the real system, one usually states that the possible health states of the believed system coincide with the possible health states of the real system, given the observations. But now some questions arise: what is a "good" representation? Is "good" a single property or a series of properties each one with a different influence on diagnostic results? What can one expect from a "good" representation? Is it possible to measure "how good" a representation is? Can one deduce how "good" a representation is based on how "good" a diagnostic result is?

The main contribution of this paper is an answer provided to all these questions. In Section 2 a framework of diagnosis inspired on the works of de Kleer, Mackworth and Reiter's in (de Kleer *et al.*, 1992) and based on the difference between the believed and the real systems is introduced. In Section 3 properties of believed systems, diagnostic problems and diagnostic results are introduced. Finally, in Section 4, a series of relations between believed system, diagnostic problem and diagnostic results properties are proved.

The importance of such relations, sometimes seen as intuitive, is that they are necessary to formally prove the accordance between the real-world system and the believed system; to formally prove that a believed system and a diagnostic problem will produce high-quality diagnostic results; or even to ease diagnostic algorithms, since for systems and problems with certain properties, different model-based diagnostic approaches produce the same diagnostic results.

[1] The believed system is also called "model" in the literature (for example in (Console and Torasso, 1991)). However, such word will not be used in this paper since it will be reserved for a model theoretic context.

2 TOWARDS A FRAMEWORK OF DIAGNOSIS DISTINGUISHING THE REAL AND BELIEVED SYSTEMS

As stated in the introductory part of this paper, this section is devoted to the formalisation of some concepts, whose understanding is almost always implicitly taken for granted in the literature about model-based diagnosis. As so, a diagnostic framework based on the difference between the believed and the real systems will be introduced. This framework will rely on first-order logic since it encapsulates almost every diagnostic approach with appealing properties in the literature.

2.1 Motivation for a framework of diagnosis distinguishing the real and believed systems through an example

Let us start by presenting an example whose objective is to motivate the need for a diagnostic framework distinguishing the believed and the real systems.

Consider the voltage divider of Figure 1 connected, for instance, to a voltage source of 9V so that the *input parameter* $V_{in} = 9V$. Suppose that one day, for instance at 11 a.m., the 21st September 2010 in Lisbon, John measures the voltage out of the voltage divider to be 6V so that the *output parameter* $V_{out} = 6V$. With his knowledge of physics John knows that in the voltage divider the relation $V_{out} = \frac{R_2}{R_1+R_2} \cdot V_{in}$ holds and he determined $\frac{R_2}{R_1+R_2} = \frac{2}{3}$. Another day, suppose for instance, at 3 a.m., the 21st December 2010 in Lisbon, for the same value of V_{in} John measured $V_{out} = 5V$, which sets $\frac{R_2}{R_1+R_2} = \frac{5}{9}$. This being the case, John could jump to the conclusion that either one or both resistors are *abnormal*, since the value differs between experiences. However, in this case he forgot another important input parameter: temperature. In fact, the resistivity of a material also depends on its temperature. As so, the different results could be explained by the material the resistors are made of, and did not indicate an abnormal resistor. In fact, in this case, no resistor was abnormal and the believed system was simply not "good" enough.

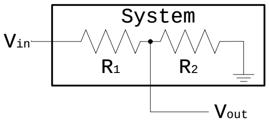

Figure 1: A simple voltage divider

This example clearly shows an interest in distinguishing the real and believed systems in a diagnostic framework; and clearly defining concepts such as, for example, *abnormality*, *input parameter* or *output parameter* before any attempt to define properties and explore relations between believed system, diagnostic problem and diagnostic results properties.

2.2 Towards a framework of diagnosis distinguishing the real and believed systems

Some concepts in this subsection have already been introduced in (de Kleer *et al.*, 1992) and (Ribot *et al.*, 2009), in which case they will be reused. The first notion to be presented is the definition of believed system (also called "model" in the literature):

Definition 1 (Believed system). *A believed system $\mathcal{S}$ is a pair $(SD, COMPS)$ where:*

1. *SD, the system description, is a set of first-order sentences.*
2. *$COMPS$, the system components, is a finite set of constants.*

Note that it is an hypothesis of this paper that there is a bijection between the real and believed system components. The notions of parameter and parameter value follow concept of believed system and correspond to the Struss' concepts of local variables and its values in a certain situation in (Struss, 1992):

Definition 2 (Parameter). *A parameter (real or believed), noted p, designates any quantity (in terms of mass or energy) that can be exchanged between a system (real or believed) and its surroundings conveying information. The set of parameters of a system is noted P_{real} or P_{bel} depending on if it is referring to the parameters of the real or believed systems respectively.*

Hereafter, the terms *real* and *believed* will be omitted by default when referring to parameters in unambiguous situations.

Definition 3 (Value of a parameter). *The value of a parameter is the value assigned to the quantity designated by the parameter. If b is a constant, $v(p) = b$ states that the value of the parameter p is b.*

The next two concepts to be presented are the real system input and output parameters (note that, in (Struss, 1992) no such distinction is made):

Definition 4 (Real system input/output parameters). *The real system parameters can be classified in two types: input parameters and output parameters.*

A parameter is an output parameter if it designates any quantity exiting the system and conveying information. The real system output parameter values can be changed by the behaviour of the real system or by the real system input parameter values.

A parameter is an input parameter if it designates any quantity entering the system and conveying information. The real system input parameter values cannot be changed by the behaviour of the real system or by the real system output parameter values.

In order to illustrate Definition 4 consider two examples. First, imagine a laptop with a keyboard and a screen. In this case, information coming from the keyboard is an input parameter of the system while the information coming out of the screen is an output parameter. No other combination of input and output parameters exists, since we cannot "write letters on the screen so that they get pressed on the keyboard". Now, consider, for example, a simple resistor as illustrated by Figure 2. In this case, it can either be the case that p_1 (the current flowing through the resistor) is an input parameter and p_2 (the voltage accross the resistor) an output parameter or that p_1 is an output parameter and p_1 an input parameter. In fact, if a current source is connected to the system, then the former case is present, while if a voltage source is connected the latter is present. What actually differentiates both situations is the *context* of utilisation. So, intuitively, a context (undefined in (Struss, 1992) since the notions of input and output parameters do not exist) is a possible structure of the system in terms of its inputs and outputs. Such concept is now formalised in Definition 5.

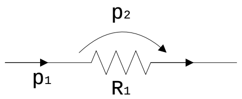

Figure 2: A simple resistor

Definition 5 (Contexts of the real system)**.** *A context of the real system, m, partitions the set of real system parameters P_{real}, into the sets of input and output real system parameters, IP_{real} and OP_{real} respectively* [2].

The set of every possible contexts of the real system, CXT_{real}, is made only of first order sentences.

Following the notion of context comes the definitions of believed system input and output parameters and believed system contexts.

Definition 6 (Believed system contexts and input/output parameters)**.** *The definitions of believed system contexts and input or output parameters is the same as Definitions 4 and 5 with the word "real" replaced by the word "believed", OP_{real} by OP_{bel}, IP_{real} by IP_{bel} and CXT_{real} replaced by CXT_{bel}. Moreover, every believed system parameter is a part of the system description SD.*

Note that we have not yet committed to any relation between the parameters or contexts of the believed and real systems. This will be a subject of discussion from here on. It is also important to note that in (Struss, 1992) it is always implicitly assumed that the believed system parameters are always a subset of the real system parameters. Finally, even the notion of real system is, at most, implicit in (Struss, 1992) since the closest concept to it, the one of "ideal" correct model of behaviour, is never defined as isomorphic to the real system as it should (shown later in this paper).

Observations are one of the few connections between real and believed systems in the diagnostic scene. Intuitively, observations are captured *from the real system* and used along with the believed system in the diagnostic reasoning (an idea shared with (Struss, 1992)). The concepts of *parameter observation* and *context observation* are now ready to be introduced:

Definition 7 (Context observation)**.** *The context observation (of a real system) is a set OBS_{real}^{CXT} of first order sentences. It represents the set of observed undistinguishable possible contexts of the real system.*

Definition 8 (Parameter observations)**.** *The parameter observations (of a real system) is a set OBS_{real}^{P} of first order sentences of the form $v(p) = b$, where p is a real system parameter and b is the observed value of p.*

Moreover, given context observation OBS_{real}^{CXT} the parameter observations OBS_{real}^{P} can be divided in those observations that are about:

- *input parameters for every context m in OBS_{real}^{CXT}, called OBS_{real}^{I}.*
- *output parameters for every context m in OBS_{real}^{CXT}, called OBS_{real}^{O}.*
- *input parameters in some contexts and output parameters in others, called OBS_{real}^{IO}.*

and $OBS_{real}^{P} = OBS_{real}^{I} \cup OBS_{real}^{O} \cup OBS_{real}^{IO}$.

Hereafter, the fact that the observations are about the real system will be omitted and the notation will be abbreviated so that OBS_{real}^{CXT} becomes OBS_{CXT} and so on. The notion of *diagnostic problem*, analogous to the notion of "abduction problem" of (Console and Torasso, 1991) and to the notion of "system" of (de Kleer *et al.*, 1992), will now be introduced:

Definition 9 (Diagnostic problem)**.** *A diagnostic problem DP is a tuple $(SD, COMPS, OBS_{CXT}, OBS_{P})$.*

We will assume all along this paper that given a diagnostic problem $DP = (SD, COMPS, OBS_{CXT}, OBS_{P})$ the parameters in OBS_{P} are a subset of P_{bel}; for there is no reason for a diagnostic system to try to observe a parameter that is not a believed system parameter.

Since there is now a *diagnostic problem* one can jump into solving it. Using de Kleer and Williams words (in (de Kleer and Williams, 1987)) solving a diagnostic problem consists of "assigning credit or blame to parts" based on observations. "Blame" and "credit" correspond to the notions of normal and abnormal which are defined, in this paper, as follows [3]:

Definition 10 (Normality and abnormality of components)**.** *A component c is said to be abnormal, noted $Ab(c)$, if it has passed its elastic limit and is deformed irreversibly.*

A component is said to be normal, noted $\neg Ab(c)$, if it is not abnormal.

The concept of abnormality is well illustrated in the example of Subsection 2.1. In that example one can understand that it is not just because the voltage divider provided a voltage output with an unusual value that the resistors are abnormal. In fact, the unusual value of the voltage output was related to temperature, an input parameter of the real system that was not taken into account in the believed system. Before providing a way for finding solutions to diagnostic problems, the notion of health state (used both in the context of real and believed systems) needs to be added.

Definition 11 (Health state)**.** *Let $\Delta \subseteq COMPS$ be a set of abnormal components. The health state $\sigma(\Delta, COMPS - \Delta)$ is the conjunction:*

$$[\textstyle\bigwedge_{c \in \Delta} Ab(c)] \wedge [\textstyle\bigwedge_{c \in (COMPS - \Delta)} \neg Ab(c)]$$

As for solving a diagnostic problem, there are two major model-based approaches:

- The first way, called in this paper the consistency-based way, consists in, for a given diagnostic problem DP, determining all the possible believed system health states that are consistent with the observations and with the system description. Then, using the fact that the believed system represents the real system, it is stated that the real system health state is exactly the same as one of the possible believed system health states. Examples of consistency-based diagnostic reasoning can be found, for instance, in (Davis, 1984), (de Kleer and Williams, 1987) or (Reiter, 1987).
- The second way is called, in this paper, the abduction-based way. For a given diagnostic

[2] In set theory, a partition of a set X is a division of X into disjoint subsets of X whose union is X.

[3] The predicate Ab is borrowed from (de Kleer *et al.*, 1992).

problem DP one starts by separating the observations in two different sets, the first one consisting of the observations to be "explained" and the second one consisting of the observations used for consistency checking. A believed system health state is then possible if it is consistent with the observations used for consistency and with the system description and if, along with the system description and with the observations used for consistency logically implies the observations to be "explained" (note that this partitioning has nothing to do with contexts as introduced in Definitions 5 and 6). Then, as before, using the fact that the believed system represents the real system, it is stated that the real system health state is exactly the same as one of the possible believed system health states. Examples of abduction-based diagnostic reasoning can be found, for instance, in (Poole, 1988) or (Console *et al.*, 1989).

In a more formal manner:

Definition 12 (Consistency-based diagnosis)**.** *Let* $\Delta \subseteq \mathit{COMPS}$. *A consistency-based diagnosis for the diagnostic problem* $\mathit{DP} = (\mathit{SD},\mathit{COMPS},\mathit{OBS}_{\mathit{CXT}},\mathit{OBS}_P)$ *is* $\sigma(\Delta,\mathit{COMPS} - \Delta)$ *such that:*

$$\mathit{SD}\cup\mathit{OBS}_{\mathit{CXT}}\cup\mathit{OBS}_P\cup\sigma(\Delta,\mathit{COMPS} - \Delta)$$

is satisfiable.

Definition 13 (Abduction-based diagnosis)**.** *Let* $\Delta \subseteq \mathit{COMPS}$. *An abduction-based diagnosis for the diagnostic problem* $\mathit{DP} = (\mathit{SD},\mathit{COMPS},\mathit{OBS}_{\mathit{CXT}},\mathit{OBS}_P)$ *with* OBS_{exp} *being the observations to be explained and* OBS_{cons} *being the observations used for consistency such that* $\mathit{OBS}_{exp}\cup\mathit{OBS}_{cons} = \mathit{OBS}_{\mathit{CXT}}\cup\mathit{OBS}_P$ *is* $\sigma(\Delta,\mathit{COMPS} - \Delta)$ *such that:*

$$\mathit{SD}\cup\mathit{OBS}_{cons}\cup\sigma(\Delta,\mathit{COMPS} - \Delta)$$

is satisfiable, and

$$\mathit{SD}\cup\mathit{OBS}_{cons}\cup\sigma(\Delta,\mathit{COMPS} - \Delta) \models \mathit{OBS}_{exp}.$$

2.3 Some words on abduction-based diagnoses

Before ending this section closer look is taken at the definition of abduction-based diagnosis that comes directly from (de Kleer *et al.*, 1992) and which also reflects other works of abduction-based diagnosis. In it, there is no formal way of choosing the sets $\mathsf{OBS}_{\mathsf{cons}}$ and $\mathsf{OBS}_{\mathsf{exp}}$ based on the sets $\mathsf{OBS}_{\mathsf{P}}$ and $\mathsf{OBS}_{\mathsf{CXT}}$. Even if in (de Kleer *et al.*, 1992) there are some "guidelines" for such task (which do not even exist in almost every work about abduction-based diagnosis), the choice of such sets is left to the common-sense of the human using the diagnostic system. Since the first goal of this paper is to leave no room for "chance" or "common-sense" in the diagnostic framework presented, we contribute with a correct way of choosing the sets $\mathsf{OBS}_{\mathsf{cons}}$ and $\mathsf{OBS}_{\mathsf{exp}}$:

Definition 14 (Correctly chosen abduction-based diagnosis)**.** *An abduction-based diagnosis for the diagnostic problem* $\mathit{DP} = (\mathit{SD},\mathit{COMPS},\mathit{OBS}_{\mathit{CXT}},\mathit{OBS}_P)$ *is said to be correctly chosen iff:*

- $\mathit{OBS}_{\mathsf{cons}} = \mathsf{OBS}_{\mathsf{I}}\cup\mathsf{OBS}_{\mathsf{IO}}\cup\mathsf{OBS}_{\mathsf{CXT}}$
- $\mathsf{OBS}_{\mathsf{exp}} = \mathsf{OBS}_{\mathsf{O}}$

Intuitively if an abduction-based diagnosis is correctly chosen then the diagnostic system will not try to explain observations about the inputs which would be a nonsense.

3 DEFINING PROPERTIES IN BELIEVED SYSTEMS, DIAGNOSTIC PROBLEMS AND DIAGNOSTIC RESULTS

In this section and, more precisely, in the six subsections that come next, eight properties related to believed systems, diagnostic problems and diagnostic results are presented, being the formalisation of some and the introduction of others our contribution.

Since the definition of some properties relies on model-theory the readers not familiar with it may refer to Appendix A or to (Hodges, 1993).

3.1 Validity and certainty of a diagnostic result

Two important notions related to a diagnostic result (the set of all possible health states of the believed system determined either in a consistency-based or in an abduction-based manner for a given diagnostic problem) are now presented: *validity* and *certainty*.

Definition 15 (Validity of a diagnostic result)**.** *A diagnostic result is said to be valid iff one of its elements is the real system health state.*

Definition 16 (Certainty of a diagnostic result)**.** *A diagnostic result is said to be certain iff it contains one and only one health state. Moreover, the more health states are in a diagnostic result, the less certain it is.*

To gain intuition about such concepts suppose a real system with two components c_1 and c_2; being, in reality, c_1 abnormal and c_2 normal. Suppose two diagnostic results: $\mathsf{DR}_1 = [\mathsf{Ab}(\mathsf{c}_1)\wedge\mathsf{Ab}(\mathsf{c}_2)]$ and $\mathsf{DR}_2 = [\mathsf{Ab}(\mathsf{c}_1)\wedge\neg\mathsf{Ab}(\mathsf{c}_2)]\vee[\neg\mathsf{Ab}(\mathsf{c}_1)\wedge\neg\mathsf{Ab}(\mathsf{c}_2)]$. DR_1 is invalid but certain and DR_2 is valid but uncertain. Finally, validity can be seen as the property guaranteed by valid models in (Struss, 1992) over the past observations; or the necessary property that works about monotonic reasoning in diagnostic such as (Ribot *et al.*, 2009) implicitly assume.

3.2 Satisfiability of the believed system and of the diagnostic problem

In Definition 1 it was stated that a believed system was a pair $(\mathsf{SD},\mathsf{COMPS})$ where SD was a first-order theory. The first property of interest is the satisfiability of a believed system:

Definition 17 (Satisfiability of a believed system)**.** *A believed system is said to be satisfiable if the theory* SD *has a model.*

This apparently unflavoured property is extremely interesting. In fact, if a believed system is not satisfiable it cannot represent the real system since the latter exists. Moreover, courtesy of Gödel completeness theorem (see (Hodges, 1993) for details), a first-order theory has a model iff it is consistent. Thus, a purely syntactic check is enough to assess this property.

Definition 18 (Satisfiability of the diagnostic problem)**.** *A diagnostic problem is said to be satisfiable if the theory* $\mathit{SD}\cup\mathit{OBS}_{\mathit{CXT}}\cup\mathit{OBS}_P$ *has a model.*

Being stronger than the satisfiability of the believed system, the satisfiability of the diagnostic problem has an added interest attached since it helps measuring the adequacy of the believed system to the real system. More precisely, if a a diagnostic problem is not satisfiable (for a satisfiable believed system), then the believed system cannot represent the real system since the real system itself and the observations about it, existing, must be satisfiable. This property will be studied with more detail in Section 4. Such concept corresponds to the "remark 1" of (de Kleer *et al.*, 1992).

3.3 Truth of the believed system

Let us contribute with a third property of interest: the truth of the believed system. It is closely related to the properties introduced in Subsection 3.2 but is much stronger. However, since there is no such thing as perfection, it is much more difficult to assess than the two properties introduced before. In fact, means will be provided to prove that a believed system is not true, but no means exist to deduce its truthfulness.

Before jumping into the definition of a true believed system it is important to understand what an axiom of a theory is. An axiom is a proposition that cannot be proved and is considered to be universally *true*, that is, true in every model of the theory. So, for instance, one can build a theory based on the axiom "All birds fly".

Another kind of truth that is different from the one introduced in the paragraph before is the notion of ontological truth. The words of Tarski provide a proper informal introduction to the subject: "a [ontological] true sentence is one which says that the state of affairs is so and so, and the state of affairs is indeed so and so" (Tarski, 1936). So, for instance, the famous sentence "All men are mortal" is an ontological truth, since all men are indeed mortal. However, the also famous sentence "All birds fly" is not an ontological truth since, in reality, there are some birds that do not fly. What is important understand is the difference between "logical truths" and "ontological truths", since the former class represents all axioms in our theory and the latter represents the correspondence between the sentences in our theory and the "real world".

The notion of ontological truth of a sentence is formally defined below:

Definition 19 (Ontological truth of a sentence)**.** *A sentence is an ontological truth iff it has a model $\mathcal{M}$ that that can be extended to a model $\mathcal{N}$ isomorphic to the "real world"* [4].

Definition 20 (Truth of the believed system)**.** *A believed system is said to be an ontological truth (or simply true) iff it has a model $\mathcal{M}$ that can be extended to a model $\mathcal{N}$ isomorphic to the "real world".*

This property is highly important; for if one wants a system description to be used in "real world" situations, then this same "real world" must always be a model of the theory. The consequences of trying to solve a diagnostic problem based on a believed system that is not true are studied in the next section.

In order to provide a visual representation of this and every other property that follows, consider the real system depicted in Figure 3

Now, imagine one of the many possible true believed systems such as the one depicted in Figure 4. Every connection between a believed system input and output parameter, if it exists, is exactly the same as in the real system. Moreover, there is no believed system parameter which is not a real system parameter. Finally, a believed system can be true even if not every real system parameter is a believed system parameter.

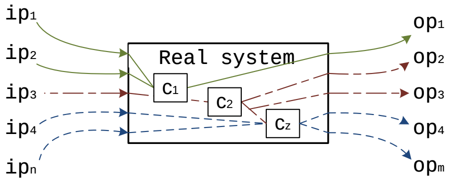

Figure 3: A typical real system

Before ending this section, it is also important to note that the truth of the believed system is not exactly the same as Struss' strong models in (Struss, 1992). In fact, true believed systems are strong models but the opposite isn't true, since true believed systems refer to the whole parameter value space and not to its domain restriction.

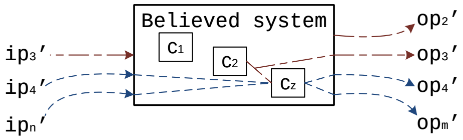

Figure 4: A true believed system

3.4 Completeness of the believed system

Up to now three properties have been introduced and we contribute in this paper with a fourth one: the completeness of the believed system.

Definition 21 (Completeness of the believed system)**.** *Let $m \in CXT_{bel}$ be a context separating P_{bel} into IP_{bel} and OP_{bel}. Let Φ_m be a set containing first order sentences of the form $v(ip) = b$ assigning a value for every $ip \in IP_{bel}$. Finally, let σ be a health state of the believed system. In this case, a believed system is said to be complete iff for every possible context m, for every different Φ_m and for every different health state σ:*

If $SD \cup \Phi_m \cup \sigma$ is satisfiable,
then $SD \cup \Phi_m \cup \sigma$ is a complete theory

Intuitively, the completeness property is a measure of how many "loose strings" there are in the believed system. In fact, if a believed system is complete one is guaranteed to know the information flow from the input parameters towards the output parameters for every context, every health state and every input parameter values. Note that even if there is a relation, completeness is very different from truth. Moreover, no such concept exists in any work known by the authors. Finally, one interesting aspect of this property is that it can, in theory, be measured; for it relies on two mathematical concepts, satisfiability and completeness of theories, that can, in theory, be proved (for instance using Vaught's test (Marker, 2002)).

Now, imagine one of the many possible complete believed systems such as the one depicted in Figure 5. Note that every connection between a believed system input and output parameter exists, even if it is not exactly the same as in the real system of Figure 3.

3.5 Coverage of the believed system

We now contribute with another property of believed systems: coverage.

Definition 22 (Coverage of the believed system)**.** *A believed system is said to fully cover the corresponding real system iff $P_{real} = P_{bel}$ and $CXT_{real} = CXT_{bel}$.*

[4] "Real world" is the "real system" and its surroundings.

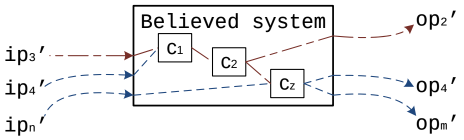

Figure 5: A complete believed system

If the completeness was said to be a measure of how many "loose strings" there are in a believed system, then the coverage can be seen as a measure of how many "strings" of the believed system are "strings" of the real system. More formally, if a believed system fully covers a real system, one is guaranteed to know in the believed system every channel of the real system from where information enters and leaves. Once again, although related, coverage and truth are very different. Although extremely attractive, this property lacks of verifiability. In fact, it is possible to prove, through the experience obtained with diagnostic results, that a believed system does not fully cover a real system, but no means exist to prove that it does cover. Finally, no such concept exists any work known by the authors.

Now, imagine one of the many possible fully covered believed systems such as the one depicted in Figure 6. Note that every real system parameter is a believed system parameter, even if the connections between parameters are not identical to the real system ones and if not every connection exists.

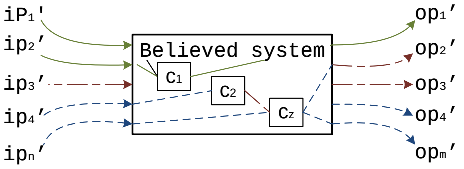

Figure 6: A fully covered believed system

3.6 Observability of the diagnostic problem

The last property of interest to us is the observability of the diagnostic problem.

Definition 23 (Observability of the diagnostic problem)**.** *A diagnostic problem $DP = (SD,COMPS,OBS_{CXT},OBS_P)$ is said to be:*

- *weakly observed iff OBS_{CXT} contains one and only one mapping and every input parameter IP_{bel} is given a value in OBS_P.*
- *strongly observed iff OBS_{CXT} contains one and only one mapping and every parameter P_{bel} is given a value in OBS_P.*

The observability of the diagnostic problem is a good measure of how much information the diagnostic system has to perform a diagnosis. In fact, for instance, given the diagnostic Definition 12, the more observed is the diagnostic problem (i.e. the more believed parameters are observed) the less possible health states are given as a solution of the consistency-based diagnostic problem. Finally, the good news is that the observability of the diagnostic problem can be measured, since one can verify how many parameters of the believed system are observed in the diagnostic problem.

Now, imagine one of the many possible strongly observed diagnostic problems such as the one depicted in Figure 7. Every believed system parameter is observed in the real system.

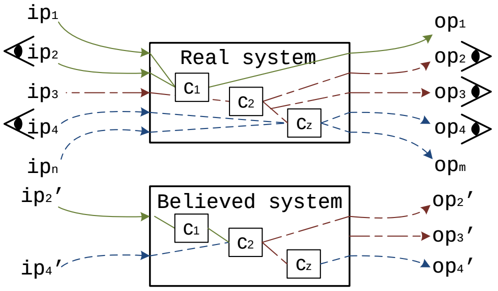

Figure 7: A strongly observed diagnostic problem

4 EXPLORING THE PARTICULARITIES OF SYSTEMS WITH GIVEN PROPERTIES

In Section 3 the properties of believed systems and diagnostic problems were introduced. It is now time to provide another contribution by explicitly deriving a series of attributes of systems that have given properties.

First of all, if a believed system is true then it is satisfiable and any diagnostic problem based on it is also satisfiable.

Lemma 1. *Every true believed system is satisfiable.*

Proof. A true believed system must have a model (cf. Definition 20), thus it is satisfiable. □

Theorem 2. *If a believed system is true, then any diagnostic problem based on it is satisfiable.*

Proof. Suppose an unsatisfiable diagnostic problem. In this case: 1) either the associated believed system is unsatisfiable or 2) the believed system is satisfiable but its union with the observations is unsatisfiable.

Case 1): In this case, by Lemma 1 one gets that the believed system is not true. Q.E.D.

Case 2): In this case the believed system has at least a model since it is satisfiable; and since the observations come from the real system, then the theory $OBS_{CXT} \cup OBS_P$ is made only of ontological truths and, as so, has at least one model whose extension is isomorphic to the "real world". Since the theory $SD \cup OBS_{CXT} \cup OBS_P$ has no model, it means, in particular, that the "real world" is not isomorphic to any extension of a model of the theory. Combining all these arguments one gets that the "real world" is not isomorphic to any extension of a model of the believed system and, as so, it is not true. Q.E.D. □

Now, that the relations that hold between the truth of a believed system and the satisfiability of believed systems and diagnostic problems have been unveiled, one can move to the relation between the truth of a believed system and the validity of the diagnostic results:

Theorem 3. *If S is a true believed system, then, for every diagnostic problem based on S:*

1. *the set of possible health states determined in a consistency-based way is valid.*
2. *the set of possible health states determined in an (correctly chosen) abduction-based way is not always valid.*

Proof. 1) Let us suppose the real state σ_{real} is not a believed system health state determined in a consistency manner. This being so, $SD \cup OBS_{CXT} \cup OBS_P \cup \sigma_{\text{real}}$ is unsatisfiable, thus having no model. Now, since

$\mathsf{OBS_{CXT}}$, $\mathsf{OBS_{P}}$ and σ_{real} come from the real system, $\mathsf{OBS_{CXT}}\cup\mathsf{OBS_{P}}\cup\sigma_{\mathsf{real}}$ must have a model that can be extended to a model isomorphic to the "real world". Combining all the arguments we get that the "real world" is not isomorphic to any extension of a believed system model. As so, it is not true. Q.E.D.

2) Suppose a true but incomplete believed system with a single output parameter whose value is never logically implied by the believed system along with the input parameters. In this case, for every σ such that $\mathsf{SD}\cup\mathsf{OBS_{cons}}\cup\sigma$ is satisfiable, $\mathsf{SD}\cup\mathsf{OBS_{cons}}\cup\sigma \not\models \mathsf{OBS_{exp}}$ and $\mathsf{SD}\cup\mathsf{OBS_{cons}}\cup\sigma \not\models \neg\, \mathsf{OBS_{exp}}$. So, every abduction-based diagnosis results in an empty set of diagnostic results, which, obviously, cannot contain the real system health state. Q.E.D. □

Theorem 3 is the actual rational for a diagnostic reasoner to be consistency-based instead of abduction-based. In fact, suppose that one "believes" [5] a believed system to have no other property than truth. In this case one is forced to choose a consistency-based reasoning over an abduction-based one, or there would be no guarantees of valid results.

The next theorem concerns the relation between consistency-based and abduction-based diagnoses for the class of complete believed systems in weakly observed diagnostic problems.

Theorem 4. *Let $\mathcal{S}$ be a complete believed system. Moreover, let DP be a diagnostic problem based on $\mathcal{S}$ at least weakly observed. In this case for every σ:*

$$\mathit{SD}\cup\mathit{OBS_{CXT}}\cup\mathit{OBS_{P}}\cup\sigma \text{ is satisfiable}$$
$$\textit{iff}$$
$$\mathit{SD}\cup\mathit{OBS_{I}}\cup\sigma \text{ is satisfiable and}$$
$$\mathit{SD}\cup\mathit{OBS_{I}}\cup\sigma \models \mathit{OBS_{O}}$$

Proof. $\Leftarrow$: If $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma$ is satisfiable, then it has a model. Since $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \mathsf{OBS_{O}}$ then, in particular, there is a model of $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma$ which is a model of $\mathsf{OBS_{O}}$. This being the case, since $\mathsf{OBS_{I}}\cup\mathsf{OBS_{O}} = \mathsf{OBS_{CXT}}\cup\mathsf{OBS_{P}}$, then $\mathsf{SD}\cup\mathsf{OBS_{CXT}}\cup\mathsf{OBS_{P}}\cup\sigma$ has a model, thus being satisfiable. Q.E.D.

$\Rightarrow$: Since $\mathsf{OBS_{I}}\cup\mathsf{OBS_{O}} = \mathsf{OBS_{CXT}}\cup\mathsf{OBS_{P}}$, $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\mathsf{OBS_{O}}\cup\sigma$ is satisfiable so one only needs to prove that $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \mathsf{OBS_{O}}$. For the sake of the argument, let φ be the sentence representing the conjunction of every element in $\mathsf{OBS_{O}}$. Since $\mathcal{S}$ is a complete system, $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma$ is a complete theory and, this being so, either $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \varphi$ or $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \neg\varphi$. Suppose $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \neg\varphi$ which is the same as saying that $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma\cup\{\varphi\}$ is unsatisfiable. However, this cannot be the case because $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma\cup\{\varphi\}$ is satisfiable. Thus, $\mathsf{SD}\cup\mathsf{OBS_{I}}\cup\sigma \models \varphi$. Q.E.D. □

So, abduction becomes consistency (and vice-versa) for a complete believed system in a at least weakly observed diagnostic problem. Once again, this relation is extremely useful due to the verifiability of completeness and observability. In fact, for a system with the discussed properties, one automatically knows that all possible health states explain the output observations.

Let us now combine Theorems 3 and 4:

Corollary 5. *If $\mathcal{S}$ is a true and complete believed system and DP is a fully observed diagnostic problem based on $\mathcal{S}$, then the set of possible health states of the believed system determined either in a consistency or (correctly chosen) abduction-based manner is valid.*

Proof. Immediate by Theorems 4 and 3. □

Means for comparing both consistency-based and abduction-based diagnostic approaches and for guaranteeing that a diagnostic approach gives a valid result have been introduced. It is now the time to study how to guarantee the certainty of a diagnostic result. First of all, if a believed system is false there is no point in assessing the certainty of a diagnostic result. This is because a false believed system can be, for example, the simple sentence $\bigwedge_{\mathsf{c}\,\in\,\mathsf{COMPS}}\mathsf{Ab(c)}$ which would always indicate a useless but certain diagnostic result.

Theorem 6. *Let $\mathcal{S}$ be a true, fully covered and complete believed system and DP a strongly observed diagnostic problem based on $\mathcal{S}$. In this case the diagnostic result $\mathcal{D}$ determined either in a consistency-based or in an abduction-based manner (using a classical logic) is valid; and there is are no other true believed systems or diagnostic problems than lead to a diagnostic result more certain than $\mathcal{D}$.*

Proof. Since Theorem 4 can be applied this proof is only consistency-based oriented.

Validity of the diagnostic result comes directly from the truth of the believed system and by Theorem 5.

Now, consistency-based diagnoses rely on the theory $\mathsf{SD}\cup\mathsf{OBS_{P}}\cup\mathsf{OBS_{CXT}}$. Call this theory $\mathcal{T}$. Suppose that $\mathcal{T}$ is an empty theory that becomes larger and larger, either with the addition of first-order sentences to SD or to $\mathsf{OBS_{P}}\cup\mathsf{OBS_{CXT}}$. Since the believed system is true, $\mathcal{T}$ will always be satisfiable by Theorem 2 and Definition 18. Moreover, these new sentences added can either make the believed system converge towards full coverage and completeness or make the diagnostic problem converge towards strong observability. At the same time, as new sentences are a part of $\mathcal{T}$ the diagnostic results become more and more certain due to the monotonicity property of classical logic. When the theory $\mathcal{T}$ corresponds to a fully covered complete believed system and to a strongly observed diagnostic problem any new true sentences added to $\mathcal{T}$ can only refer to entities that have no relation to the system, thus having no effect on the diagnostic results. This being so, the limit of certainty is reached with a fully covered complete believed system and a strongly observed diagnostic problem. □

5 DISCUSSION

From Theorems 5 and 6 we are able to understand that the truth of a believed system is the main driver for obtaining valid diagnostic results. Similarly, coverage and completeness (assuming truthfulness) of a believed system and the observability of the diagnostic problem are the main drivers for obtaining a certain diagnostic result. Moreover, the observability of the diagnostic problem is a main condition for finding properties in diagnostic systems.

From the application of *modus tollens* to every theorem one can actually try to deduce some properties of a certain believed system from the diagnostic problems and the diagnostic results they provide. For example,

[5] The usage of the word believe in the sentence is extremely important; for one cannot "know", that is, prove, that a believed system is true.

if a diagnostic result obtained in a consistency-based manner is different from a diagnostic result obtained in an abduction-based manner for the same believed system, then either the believed system is incomplete or the diagnostic problem is not at least weakly observed. Since the latter is easily checked, one gets an indicator on the completeness of believed systems.

What is interesting to note is that the validity and certainty of diagnostic results, the satisfiability, completeness, truth and coverage of believed systems and the satisfiability and observability of diagnostic problems are strongly related. Thus, the knowledge of some of these characteristics can be used to deduce the others as depicted in Figure 8.

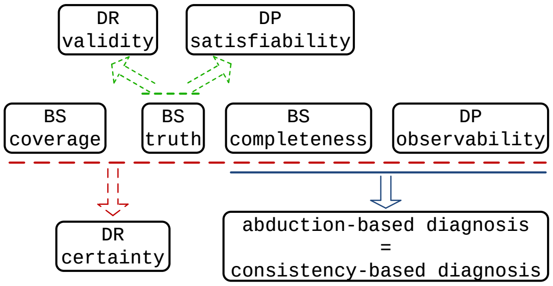

Figure 8: Believed system (BS), diagnostic problem (DP) and diagnostic results (DR) property relations

This being the case, these relations are necessary to formally prove the accordance between the real-world system and the believed system; to prove that a believed system and a diagnostic problem will produce valid and certain diagnostic results; or even to ease diagnostic algorithms in some situation where a consistency-base approach is guaranteed to provide the same results as an abduction-based approach.

REFERENCES

(Console and Torasso, 1991) Luca Console and Pietro Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 7:133–141, 1991.

(Console *et al.*, 1989) Luca Console, Daniele Theseider Dupre, and Pietro Torasso. A theory of diagnosis for incomplete causal models. In *Proc. 11th IJCAI*, pages 1311–1317, 1989.

(Davis, 1984) Randall Davis. Diagnostic reasoning based on structure and behavior. *Artif. Intell.*, 24(1-3):347–410, 1984.

(de Kleer and Williams, 1987) Johan de Kleer and Brian C. Williams. Diagnosing multiple faults. *Artif. Intell.*, 32(1):97–130, 1987.

(de Kleer *et al.*, 1992) Johan de Kleer, Alan K. Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artif. Intell.*, 56(2-3), 1992.

(Hodges, 1993) Wilfrid Hodges. *Model Theory*. Number 42 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1993.

(Marker, 2002) David Marker. *Model Theory: An Introduction*, volume 217 of *GTM*. Springer-Verlag, New York, NY, 2002.

(Poole, 1988) David Poole. Representing knowledge for logic-based diagnosis. In *FGCS*, 1988.

(Reiter, 1987) Raymond Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32(1), 1987.

(Ribot *et al.*, 2009) Pauline Ribot, Yannick Pencolé, and Michel Combacau. Diagnosis and prognosis for the maintenance of complex systems. In *Proc. SMC'09, IEEE International Conference on Systems, Man, and Cybernetics*, 2009.

(Struss, 1992) Peter Struss. *What's in SD?: Towards a theory of modeling for diagnosis*, pages 419–449. Morgan Kaufmann Publishers Inc., 1992.

(Tarski, 1936) Alfred Tarski. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics*, pages 152–278. Oxford University Press, Oxford, 1936.

A SOME WORDS ON MODEL-THEORY

In this paper we use the notions of: model, isomorphism, extension, theory and complete theory. These concepts are clarified hereafter. The material presented is based on (Hodges, 1993) and (Marker, 2002).

First of all, a structure $\mathcal{M}$ is an object specified by:

- A (possible empty) set called the domain or universe of $\mathcal{M}$ written M.
- A (possible empty) set of constant elements of $\mathcal{M}$, each named by one or more constants.
- For each positive integer n, a (possible empty) set of n-ary relations on M. Each relation is named by one or more n-ary relation symbols.
- For each positive integer n, a (possible empty) set of n-ary operations on M. Each operation is named by one or more n-ary function symbols.

The signature $\mathcal{L}$ of the structure $\mathcal{M}$ is specified by the set of constants, relation symbols and function symbols of $\mathcal{M}$ (we assume $\mathcal{L}$ can be read off uniquely from the structure). We will also use the symbol $\mathcal{L}$ to indicate the language generated by the signature $\mathcal{L}$.

Let $\mathcal{M}$ and $\mathcal{N}$ be two $\mathcal{L}$-structures with domains M and N respectively. An $\mathcal{L}$-embedding $\mathsf{f}\colon \mathcal{M} \to \mathcal{N}$ is a one-to-one map $\mathsf{f}\colon \mathsf{M} \to \mathsf{N}$ that preserves the interpretation of all the symbols of $\mathcal{L}$. A bijective $\mathcal{L}$-embedding is an $\mathcal{L}$-isomorphism.

Now, if $\mathsf{M} \subseteq \mathsf{N}$ and the inclusion map is an $\mathcal{L}$-embedding, we say that $\mathcal{N}$ is an extension of $\mathcal{M}$ or that $\mathcal{M}$ is a substructure of $\mathcal{N}$.

The language $\mathcal{L}$ consists of formulas, that is, strings of symbols built using rules of grammar, the symbols of the signature $\mathcal{L}$, variable symbols, the equality symbol, the boolean connectives, quantifiers and parentheses. Moreover, a theory is simply a set of sentences.

To make a long story short, if $\mathcal{M}$ is a $\mathcal{L}$-structure, then each $\mathcal{L}$-sentence φ is either true or false in $\mathcal{M}$. If φ is true in $\mathcal{M}$, $\mathcal{M}$ is said to be a model of φ and noted as $\mathcal{M} \models \varphi$. Moreover, given a theory T, $\mathcal{M}$ is said to be a model of T if $\mathcal{M} \models \varphi$ for all sentences $\varphi \in \mathsf{T}$. This is written as $\mathcal{M} \models \mathsf{T}$. A theory with at least a model is said to be satisfiable.

Given a theory T and a sentence φ, φ is said to be a logical consequence of T, written $\mathsf{T} \models \varphi$, if $\mathcal{M} \models \varphi$ whenever $\mathcal{M} \models \mathsf{T}$.

Finally, a theory T is said to be complete iff for every sentence φ in the language $\mathcal{L}$, either $\mathsf{T} \models \varphi$ or $\mathsf{T} \models \neg\varphi$