# Causal Structures: From Logical Models to Diagnostic Applications


Pilar Fuster-Parra* and Antoni Ligeza**[1]

*University of Balearic Islands, c. Valldemossa, km.7.5; 07071 Palma Mallorca, Spain

**Institute of Automatics AGH, al. Mickiewcza 30; 30-059 Krakow, Poland

fax: *+34 971 17 30 03 - e-mail: * pilar@ipc4.uib.es, ** ali@ia.agh.edu.pl



**Abstract**

The phenomenon of *causality* is omnipresent in language, philosophy and science and it seems to be a core, fundamental idea in common-sense thinking. In this paper the nature of causality is investigated and some characteristics of this concept are presented. Particular attention is also paid to applications in diagnostic reasoning. Causality provides bases for diagnostic inference and explanation of failures through search in the direction against causal influences. From logical point of view, this can be considered as *abductive inference*. A core, generic model of the search space is built as an AND/OR/NOT causal graph specifying the structure that is used to establish diagnoses. Such a graph constitutes a schematic representation of causal influences among symptoms and incorporates logic-like functions for combining influences of multiple symptoms. An illustrative example is presented to clarify most of the presented ideas.



**Keywords:** Causality, Causal Reasoning, Causal Graph, Abductive Reasoning, Backward Search, Qualitative Probabilities, Fuzzy Faults, Tabular Systems.


## 1 Introduction

One of the biggest problems concerning the so-called *model-based approaches* to diagnosis is that they tend to be very inefficient from computational point of view. Further, they require that a precise model of system behaviour must be known. Model-based approaches operate on a declarative description (of the structure and function behaviour) of the system to be diagnosed (rather than on a specialised representation suitable for diagnosis, as, for example, expert systems). However, diagnostic inference can often be performed without knowledge of a complete system model. Because of that, it is necessary to investigate approaches able to enhance the performance of model-based systems operating on diagnosis-oriented, simplified models, covering system behaviour in a *qualitative* way mostly(see Fuster-Parra (1996)).


[1] The research has been carried out within the KBN Grant no. 8 T11C 017 19 with respect to the second author.

In order to enable diagnostic inference, a structure called an *AND/OR/NOT causal graph* is used throughout this paper[2]. Such a graph constitutes a schematic representation of causal influences among symptoms and incorporates *logic-like functions* for combining influences of multiple symptoms. Different kinds of causal relations can be considered according to the domain of work. In this proposal we point out only to three basic types of causal influences, which explain the causal relations between symptoms in the AND/OR/NOT causal graph designed for abductive diagnostic inference. The three different causal relations refer to the 'strength' of causality and level of abstraction of the represented knowledge, according to the domain of work and the elicited knowledge.

The main idea of the proposed approach is to provide a relatively simple, straightforward 'formalism' for diagnosis, easily understandable by domain experts, based on well-studied techniques of AI. The principal inference model constitutes a core, generic representation of causal dependencies and allows for certain modifications and extensions. On informal descriptions of real-world phenomena, statements of the form 'A causes B' are exceedingly common. Causal reasoning is used to identify the possible causes of a failure in a process. In spite of abundant literature concerning causality, causal reasoning, causal networks and causal graphs, it is not quite obvious what a causal relation is, how can it be characterised, modelled and what are its formal properties.

The approaches based on causal models are attractive from model-based diagnosis point of view because of the ability to deal with incomplete and partial (uncertain) knowledge and because they need relatively small amount of information for performing diagnoses. A very attractive feature of causal approaches is their easy, intuitive interpretation and understanding as well as closed position to the engineering way of thinking.

Abductive reasoning seems to constitute a natural and straightforward diagnostic approach. However, pure logical abduction does not appear to be a valid inference rule; further, there seems to be no unique, formal approach to abductive diagnostic reasoning. One of the main points of this paper is to consider abduction as a *backward search procedure*, and the main problem is to define the *search space* and efficient search rules. As heuristic search methods constitute a well-explored domain of AI, search algorithms viewed as abductive inference strategies can be used to assure diagnostic efficiency (an extension of it can be seen in Fuster-Parra (1996)).

Some new extensions include introduction of diagnoses covering not only component faults but 'wrong' compositions of control actions and operational conditions, *fuzzy* characterisation of a degree to which components are faulty, and finite, multiple state component model. The proposed enhancements of diagnostic reasoning are related to the introduction of a formal test definition and selection criteria for sequential testing

[2] This paper is a kind of review and survey paper based on the Authors' research in the domain of automated diagnosis with use of causal reasoning; for many detailed results and a full list of papers see Fuster-Parra (1996) and Fuster-Parra and Ligeza (2000).

during diagnostic reasoning, ordering of alternative symptoms with use of *qualitative probabilities*, and a strategy for final validation of diagnosis.

A set of tools that helps to support graphical design and manipulation of causal structures has been implemented. A Prolog program is used as an inference engine for processing the symbolic structure that represents the causal graph. There is also a program in C++ that provides a graphical user interface.

## 2 The nature of causality

On informal descriptions of real-world phenomena, statements of the form '*A causes B*' are exceedingly common. Several researches have been working on Causal Reasoning, among them Iwasaki and Simon (1988), de Kleer and Brown (1986), Pearl (1988), Kuipers (1984), etc.; they propose different procedures for propagating causal disturbances. Quite a lot of researchers working on diagnosis have used 'causality' to reduce the search space during reasoning process, among them Console and Torasso (1992), etc.

In this paper the main interest is in using causal reasoning to identify the possible causes of a failure in a process. The methods based on graphs are attractive because they need relatively little information to establish a diagnosis. An example of this is the work by Kramer (1987); he uses a *signed directed graph* that represents causality ways in the failure process. The nodes are state variables and alarm conditions, the arcs are causal influences between the nodes. This graph is translated into a set of logical rules that give a framework to deal with the resolution problem of diagnosis. The main difference with the proposal presented in this paper is that the graph structure representing the causal behaviour can be used itself for searching diagnosis without translation into logical rules. The causal ordering by Iwasaki (1989) and the mythical causality by de Kleer (1986) constitute another way of using causality, different of this approach.

### 2.1 Causal relations

It appears difficult to establish a general definition about what a *causal relation* is. The postulate of a principle of causality of Gaines (1976), 'to every effect there is a cause' has been a continuing to appear as a central problem for philosophy. As some formal statement coherent with the rest of this proposal we shall establish the following general definition:

**Definition.** *A causal relation between symptoms n and n' (represented as graph nodes) is an influence of the fact that n occurs whenever n' will occur as well.*

Note that symptoms can be considered as variables taking the values *True* or *False*; from logical point of view they are equivalent to *propositional formulae*. An extended formulation can assume symptoms to be variables taking specific values (e.g. of the form $V=t$) or variables restricted to certain sets or intervals (e.g. $V>t$). In general, symptoms can be considered also as facts being expressed in some logic, e.g. first order calculus. For simplicity, the basic, generic presentation of this paper refers to symptoms mostly as propositional statements.

By the term *influence* we mean several possibilities, i.e. there is a relationship between nodes of causal type, but this relationship can be stronger or weaker depending of the knowledge. For intuition, node $n$ has influence on none $n'$ iff the occurrence of $n$ causes $n'$ to occur.

Let us turn to the causal relations between symptoms/nodes. It is assumed that whenever there is a causal relation between nodes $n$ and $n'$, there is a directed arc pointing from $n$ to $n'$. In the basic problem formulation we do not distinguish different types (or functions) of this influence; for keeping the considerations a simple arc pointing from $n$ to $n'$ says that *n causes n'* to occur or *n may cause n'* to occur. Of course, more than one symptom can influence some other symptom; in such case we have to specify the way several input symptoms influence the output one. This is done by assuming the semantics of causality and defining the causal relations.

For the purpose of this paper we define logical semantics of causality. Let us consider symptoms to be propositional formulae of some language $\mathcal{L}$. Let **I** denote some assumed interpretation, or a set of assumed interpretations (an interpretation partially specified). Further, let us consider two symptoms, say $p$ and $q$. Let $t_x$ denote the time instant when the value of symptom $x$ changes from false to true and $t_{\underline{x}}$ denote the instant of time when the value of symptom $x$ changes from *true* to *false*.

**Definition** *We shall say that symptom p has a causal influence on symptom q if and only if the following conditions hold:*

1. *At least one of the following implications holds under the assumed set of interpretations:*

   *a1)* $\models_I p \rightarrow q$,

   *a2)* $\models_I p \rightarrow \neg q$,

   *a3)* $\models_I \neg p \rightarrow q$,

   *a4)* $\models_I \neg p \rightarrow \neg q$.

   *For the sake of consistency, neither a1) and a2) nor a3) and a4) can hold simultaneously.*

2. *For the holding implications there is respectively:*

   *b1)* $t_p < t_q$,

   *b2)* $t_p < t_{\underline{q}}$,

   *b3)* $t_{\underline{p}} < t_q$,

   *b4)* $t_{\underline{p}} < t_{\underline{q}}$,

3. *There is a flow of physical signal from p to q.*

This definition is rather a strong one; in fact it refers to *strong causality*. For various purposes it can be weakened in several ways. However, for simplicity and clarity of the discussion here we keep the notion of causality to be understood as defined above. In order to keep consistency with intuition it is also assumed that *a1)* can hold with *a4)* only , while *a2)* can hold simultaneously with *a3)*.

Now we can pass to the definition of causal relation itself. With respect to the above definition, a *causal relation* is understood as an influence described further as a function or a partial function from the set of possible combinations of values of the input symptom to the values of the output symptom. Let $n_1, n_2, \ldots, n_k$ and n be symptoms; a causal relation defining an influence of $n_1, n_2, \ldots, n_k$ on n is any function of the form:

$$\Psi: (n_1, n_2, \ldots, n_k) \rightarrow n.$$

The function can be total or partial. Selection of specific functions is a matter of current requirements. In general, if the symptoms are considered to be variables taking different values, the function can be specified with a table of the form:

| *n1* | *n2* | .... | *nk* | *n* |
|---|---|---|---|---|
| *t11* | *t1* | ... | *t1k* | *v1* |
| *t21* | *t22* | ... | *t2k* | *v2* |
| ... | ... | ... | ... | ... |
| *tm1* | *tm2* | ... | *tmk* | *vm* |

**Table 1:** A tabular specification of a discrete causal function.

In case of two-valued, logical symptoms, a reasonable and straightforward choice is to take functions describing the basic logical connectives, i.e. AND, OR, NOT. For example, in case of three-input AND and OR tables the specification is as follows:

| *n1* | *n2* | *n3* | *n* |
|---|---|---|---|
| *1* | *1* | *1* | *1* |
| *0* | _ | _ | *0* |
| _ | *0* | _ | *0* |
| _ | _ | *0* | *0* |

**Table 2:** A tabular specification of the three input AND function.

| *n1* | *n2* | *n3* | *n* |
|---|---|---|---|
| *0* | *0* | *0* | *0* |
| *1* | _ | _ | *1* |
| _ | *1* | _ | *1* |
| _ | _ | *1* | *1* |

**Table 3:** A tabular specification of the three input OR function.

In the above tabular systems *1* stays for *True* and *0* for *False*; the underscore ("_") means any value. The specification of the NOT function is obvious. Note that in the case of abductive inference, two or more rows with the same value in the rightmost column are the source of branching – this value can be achieved in several ways.

### 2.2 AND/OR/NOT causal graphs

A causal graph is a structure representing all the considered causal dependencies in the analysed system. Let **N** be the set of considered symptoms and let $\Psi$ denote a set of specific functions defining causal relations among the symptoms of **N**. The causal graph is defined as follows:

**Definition.** *If N is a set of symptoms and* $\Psi$ *is a set of functions defined on these symptoms, then a causal graph is a structure* $\mathbf{G} = (N, \Psi)$.

In the graph, the nodes such that no arc points to them, will be referred to as *initial nodes*; for simplicity these nodes are denoted by **D**. The elements of D denote in fact *elementary diagnoses* (which can take the *True* or *False* values). Further, the nodes from which no arc points to other nodes will be referred to as *terminal* or *final nodes*; they are denoted by **M**. Such symptoms are also called *manifestations*. To simplify the discussion we also assume that there are no loops in the graph. All other nodes are *intermediate* ones; they are denoted with **V**. The state of the graph is defined by an assignment of the truth values to its nodes.

The defined above AND/OR/NOT causal graph is somewhat similar in structure to classical AND/OR graphs used in problem-solving by Nilsson (1971) and Pearl (1985); thus we attempt to follow the existing terminology if possible. The main difference lies in the direction and interpretation of arcs. A visible extension consists of admitting the NOT links. Further, a 'solution graph' in classical problem-solving constitutes only a 'possible justification' (to be further validated) for an observed failure.

### 2.3 Abductive search approach

A diagnostic problem exists if at least one fault is observed. The faults to be diagnosed are assumed to be specified with some manifestations (either positive or negative ones). Consider an AND/OR/NOT causal graph $G$. Formulation of diagnostic problem takes also into account some possible observations providing further information to the diagnostic system. The rules of propagation are defined below:

**Forward propagation**: (causality, simple logical interpretation assumed)

- *OR node true*: if at least one of the predecessors of an OR node is *true*, then the value of the OR node is set to *true*,
- *AND node false*: if at least one of the predecessor of an AND node is *false*, then the value of the AND node is set to *false*,
- *NOT node true*: if a predecessor of a NOT node is false, then the value of the NOT node is set to *true*,
- *NOT node false*: if a predecessor of a NOT node is *true*, then the value of the NOT node is set to *false*.

**Forward propagation**: (causality, simple logical interpretation assumed; further, *completeness* is assumed, i.e. the predecessors of a node are all the direct causes for it)

- *OR node false*: if all the predecessors of an OR node are *false*, then the value of this OR node is set to *false*,
- *AND node true*: if all the predecessors of an AND node are *true*, then the value of this AND node is set to *true*.

**Backward propagation**: (causality, simple logical interpretation assumed)

- *OR node false*: if an OR node is *false*, then the values of all its predecessors are set to *false*,
- *AND node true*: if an AND node is *true*, then the values of all its predecessors are set to *true*,
- *NOT node true*: if a NOT node is *true*, then the value of its predecessor is set to *false*,
- *NOT node false*: if a NOT node is *false*, then the value of its predecessor is set to true.

**Backward propagation**: (causality, simple logical interpretation assumed; further, *completeness* is assumed, i.e. the predecessors of a node are all the direct causes for it)

- *OR node true*: if an OR node is *true*, then at least one of its predecessors must be true (see the above OR table for an example),
- *AND node false*: if an AND node is *false*, then at least one of its predecessors must be *false* (see the above AND table for an example).

The above rules define the principles of state propagation. In fact the last two rules define the branching possibilities for search; all the other rules are deterministic ones. Whenever a rule is applicable, a new symptom value is generated; it is next placed in the set representing the current state. In case some symptom turns out to take two inconsistent values, the initial state for propagation is considered to be inconsistent and it is not taken into account any more. In practice, this has the effect of *failure* and *backtracking* in Prolog. All the calculated diagnoses are potential explanations of the observed misbehaviour; they must be further verified.

Two approaches can be differentiated: one oriented towards calculating all the potential diagnoses, and another one oriented towards calculating a first possible diagnosis, perhaps the most plausible one. Then such a diagnosis is verified, and only if it is not the correct one, the search is continued.

## 3 Types of causality

After Console and Torasso (1992), we would like to point out some extended possibilities of causal influence. They present different kinds of causal relations according to the mechanical domain of work: cause-effect, HAM (Has As a Manifestation), defined-as, must, may, ... . In this proposal we point out only to three basic types of causal influences, which explain the causal relations between symptoms in the AND/OR/NOT graph. The distinction is important from the point of view of causal diagnostic reasoning.

The three different causal relations refer to the 'strength' of causality and the level of abstraction of the represented knowledge. They also concern the problem of incompleteness of our knowledge and knowledge representation. We distinguish the three following levels of causality:

1. Symptom n causes symptom n' always when the former occurs; moreover, the occurrence of n' is bound to be caused by n. We refer to this type of causation as sufficient and necessary (NEC), i.e. we have a complete, single-cause model.
2. Symptom n causes symptom n' always when the former occurs, but there are several possible different symptoms causing n' as well. We shall refer to this type of causation as sufficient (SUF), i.e. we have a complete, many-causes model.
3. Occurrence of symptom n *may cause* symptom n' to occur, however there are cases when n' does not follow n. We refer to this weakest type of causation as *possible* (of the type MAY in Console (1992)), i.e. we have probably incomplete, many-causes model, where incompleteness is dealt with by probability.

The NEC relationship is the strongest one, it let us know a more precise information, but we must assume that in real problems it is difficult to obtain information as precise as that. We normally find different kinds of causal influence. For example 'drinking alcohol MAY cause stomach disturbances', ' rain is SUFicient to cause reduction of hydraulic stress in the crops' (it depends also on other factors, as artificial watering can cause the same effects).

In our further considerations we typically assume incompleteness of the causal knowledge, and, as a consequence of that the third type of causality will be mostly taken into account. However, profiting from additional knowledge the proposed search mechanism can be made more efficient, and such a modification seems to be useful if the appropriate knowledge is given.

### 3.1 Fuzzy faults

In this section an extension of diagnostic reasoning concerning 'degree of faultiness' (see Fuster-Parra (1996)) is included.

Most of the current approaches admit only binary evaluation of faults, i.e. a component can be faulty or correct. However, for more complex systems incorporating large number of diversified components (e.g. pneumatic, hydraulic, mechanical, chemical, biochemical, etc.) the above approach is not necessarily correct. The point is that in such systems certain components can be regarded as 'faulty to a certain degree', not just *faulty* or *correct*. The sources of such 'partial' faults usually lies in the characteristics of the components –they perform *continuous* processes with some boundary conditions determined by the state of elementary parts of them. This type of processes include ones based on different sorts of flows, concentration, exchange of energy, etc., where the state parameters change in a continuous way. After some time of working (as a natural process) or due to some accidental changes the process, may be considered in a state not fully correct (<100%).

Let us assume that the fuzzy coefficients of fault occurrence for elements of **D** (the nodes in the lowest level) are assigned to the considered elements. In order to perform operations to propagate the fuzzy measures through the causal graph the following operations are introduced:

1). The *min* and *max* operations,

2). Any *T-norm* and *S-norm* (*T-conorm*)

3). The use of triggering function.

The simplest case is the propagation of fuzzy measures through the graph using the *min*, and *max* operations. The second case is more general. Instead of propagating the fuzzy measures through the graph using *min* and *max* operations, *T-norms* and *S-norms* can be applied. The third case is the most general one. Following this line of reasoning, we can further assume that also the 'degree' or 'strength' of influence of certain fuzzy faults can be characterised with fuzzy coefficients; thus, we assume that the arcs of the causal graph are assigned some numbers $\beta_i$ where $\beta_i$ belongs to [0,1], or more generally some kind of triggering function.

In real world problems we may have fuzzy characterisation of faults. The presented model can be used in different ways.

1 Binary search/fuzzy ordering: the search for diagnoses can be performed assuming the classical binary model. In this way preference among diagnoses can be established.

2 Fuzzy model: under the assumption that the degree of faultiness is evaluable, this approach can be used to order the search for diagnoses by propagating the expected fuzzy values of faults, most likely symptoms can be searched first.

3 Simulation: the whole model can be used for simulation of influence of fuzzy faults of system components on the observed failures and estimation of expected fuzzy values of them.

### 3.2 Introducing possibility measures in an AND/OR/NOT causal graph

In this section we shall introduce the concept of *possibility measure*, which was initially presented by Zadeh (1978) and used in knowledge representation by Dubois and Prade (1985), for reasoning in an AND/OR/NOT causal graph that includes *may* connections between nodes. Whenever two nodes $n_i$ and $n_j$ are connected by a *may* arc, it is established that the occurrence of a symptom $n_i$ may cause the occurrence of another symptom $n_j$.

**Definition.** *It is said that two connected nodes ($n_i$, $n_j$) have a possibility measure associated whenever they are connected with a 'MAY' arc.*

Let **G** be an AND/OR/NOT causal graph. Assume that the possibility measures assigned to arcs *may* are known, then we assign to the other arcs (*suf* and *nec*) a possibility measure equal to 1. The following rules define how to propagate the possibility measures of symptoms upward a graph **G**:

1. The case of an OR node: let ($n_1$, $n_i$), ..., ($n_{i-1}$, $n_i$) be nodes connected by an OR-arc, and let $m_j$ denote the possibility measure assigned to $n_j$, j=1, 2, ..., i-1; then the possibility measure $m_i$ of $n_i$ is calculated as $m_i = \max(m_1, m_2, \ldots, m_{i-1})$;

2. The case of an AND node: let ($n_1$, $n_2$, ..., $n_{i-1}$, $n_i$) be nodes connected by an AND-arc, and let $m_j$ denote the possibility measure assigned to $n_j$, j=1, 2, ..., i-1; then the possibility measure $m_i$ of $n_i$ is calculated as $m_i = \min(m_1, m_2, \ldots, m_{i-1})$;

3. The case of a NOT arc: let ($n_1$, $n_2$) be nodes connected by a NOT-arc, and let $m_1$ denote the possibility measure assigned to $n_1$; then the possibility measure $m_2$ of $n_2$ is calculated as $m_2 = 1 - m_1$.

A diagnosis is understood as a conjunction of its components as it is shown in Fuster-Parra and Ligeza (1996) (all the elementary diagnoses constituting the elements of a diagnose must be observed); the possibility measure associated to a diagnosis $D=[d_1, \ldots, d_m]$ is $m= \min(m_1, \ldots, m_m)$ where $m_1, \ldots, m_m$ are the possibility measures associated to $d_1, \ldots, d_m$ respectively.

The possibility measures, which have been defined, are used for ordering the search, i.e. searching first those hypothesis with higher possibility measures. As a consequence, those diagnoses with higher possibility measure associated would be found first.

### 3.3 Qualitative probabilities and may connections

In order to get search efficiency in diagnostic reasoning process based on abductive analysis of causal structures the concept of *qualitative probabilities* of Fuster-Parra and Ligeza (1995,1996) was presented. The qualitative probabilities $Q = [q_1, \ldots, q_n]$ denote linguistic (e.g. expert-provided) statements concerning the likelihood of certain events.

An example of such statement can be $q_i = very\_likely$. Further let $\gtrsim$ denote a (weak) order relation, and let there be $q_n \gtrsim q_{n-1} \gtrsim \ldots \gtrsim q_1$. The weak order relation $\gtrsim$ means that a greater element denotes greater likelihood. The decision-maker has to have an inherent feeling of relative likelihood. For intuition, the above items denote linguistic (e.g. expert-provided) statements concerning the likelihood of certain events.

It is assumed that the qualitative probabilities of fault occurrence for elements of D are known. In the case where may connections were taken into account then a function that translates the possibilities measures into qualitative values is considered.

**Definition.** *Let $(n_i,n_j)$ be connected by a may connection, then a function $\Delta: P \rightarrow Q$ that assigns to every possibility measure a qualitative probability is called a transfer function from quantitative to qualitative knowledge representation.*

Whenever there were a may connection between two nodes $(n_i, q_i)$ and $(n_j, q_j)$ then the qualitative probability assigned to $n_i$ will be $q_i' = \min(q_i, \Delta(p_k))$. The following rules define the possibility to propagate the qualitative probabilities of symptoms upward a graph **G** in Fuster-Parra and Ligeza (1995):

1. The case of an OR node: let $(n_1, n_i), \ldots, (n_{i-1}, n_i)$ be nodes connected by an OR-arc, and let $q_j$ denote the qualitative probability assigned to $n_j$, $j=1, 2, \ldots, i-1$; then the qualitative probability $q_i$ of $n_i$ is calculated as $q_i = \max(q_1, q_2, \ldots, q_{i-1})$;
2. The case of an AND node: let $(n_1, n_2, \ldots, n_{i-1}, n_i)$ be nodes connected by an AND-arc, and let $q_j$ denote the qualitative probability assigned to $n_j$, $j=1, 2, \ldots, i-1$; then the qualitative probability $q_i$ of $n_i$ is calculated as $q_i = \min(q_1, q_2, \ldots, q_{i-1})$;
3. The case of a NOT arc: let $(n_1, n_2)$ be nodes connected by a NOT-arc, and let $q_1$ denote the possibility measure assigned to $n_1$; then the qualitative probability $q_2$ of $n_2$ is calculated as $q_2 = 1 - q_1$.

According to the above rules one can inductively assign qualitative probabilities to a maximal subset of **N**, provided that expert derived qualitative probabilities have been assigned to all elementary diagnoses (the elements of D).

From this qualitative probability ordering we can say that $D_1$ and $D_2$ being two diagnoses, and the qualitative probability associated to them are respectively $q_1$ and $q_2$ such that $q_2 \gtrsim q_1$ (and $q_2 \neq q_1$), then $D_2$ is found first as it is shown in Fuster-Parra (1996).

As a consequence if $D_1, \ldots, D_k$ are all the possible diagnoses and q = $\max_i$[qualitative probabilities($D_i$)], then D = [$D_j$ such that qualitative probability ($D_j$) = q] is found first; it is also shown in Fuster-Parra (1996).

## 4 A simple diagnostic example

### 4.1. Diagnostic Problem

In this section a simple example of a diagnostic problem and its solutions are presented. The diagnostic problem concerns a small system (possibly a part of a bigger one) composed of a tank, water supply system controlled with a valve, water-removing system driven by a pump and a level sensor. The scheme of the system is shown on the figure below.

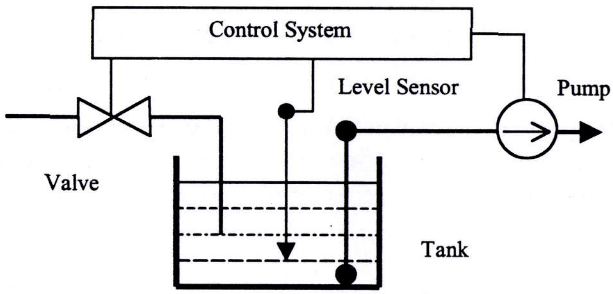


**Fig.1**: A simple system to be diagnosed.

The basic analysis of the system behaviour is as follows. The tank should normally contain certain amount of water provided through the valve. If the level is too low the valve can be opened and water flows into the tank. Its level is controlled with a single level sensor providing the possibility to close the valve after the required level is achieved. The pump is normally used to remove the water after finishing the operation of the system. If by some accident the level is too high, the pump removing the water is also started. It is assumed that both the valve and the pump are controlled (according to some control algorithm) with use of signals from level sensor and from the control system. For the sake of security the system is made safe by double protection system; whenever the level of water is too high, not only the valve should be closed, but the pump should be switched on, as well. Further, we assume that the capacity of the pump is greater than the one of the water-supplying valve.

One can consider two kinds of expected behaviour of the system with respect to the water level. The *normal expected behaviour* of the system is to keep the level of the water within some predefined limits. The *abnormal expected behaviour* may consist in an overflow, i.e. a situation when the level is too high. Let us consider the latter case, i.e. we are interested with defining, detecting and diagnosing the explicit fault rather than any abnormality occurring. The fault of interest is just one, i.e. *water overflow*.

### 4.2 Failure Recognition

Assume that at a certain moment an overflow occurs - this is an evident failure, so a diagnostic procedure is started.

For illustration, let us assume that the partial state representation is a logic-like formula of the form *water_level* $\in$ *(low,high)* for normal operation, where the meaning of *water_level* $\in$ *(low,high)* is that the level of the water is higher than some level denoted with *low* and lower than the one denoted with *high*.

Now let us consider a formula describing all the states (a situation in our terminology, i.e. a set of states) in which overflow occurs. The formula may be of the form *water_level* $\in$ *(max, +∞)*.

Let the current state formula be *water_level* $\in$ *(very_high, +∞)*, where $\phi$ denotes the rest of the facts true in the particular state. Thus in our case there is:

$$water\_level \in (max, +\infty) \geq water\_level \in (very\_high, +\infty) \wedge \phi$$

In fact, in any state described with the right hand formula above, the failure characteristics specified by the formula *water_level* $\in$ *(max, +∞)* must hold; the crucial element of the check consists in verifying that $(max, +\infty) \supseteq (very\_high, +\infty)$.

### 4.3 Symptoms and causal graph

The diagnostic process is based on causal abductive reasoning with use of a model represented by a *causal graph*. Such a graph represents causal dependencies among *symptoms* observed in the analysed system. In order to build a causal graph one should identify the set of symptoms concerning the behaviour of the system and determine the causal relationship among them.

**Manifestations, here single failure symptom:**

m -- water pouring out of the tank,

**Intermediate symptoms:**

v1 - valve_open,
v2 - pump_off,
v3 - valve_stuck_in_open_position,
v4 - valve_open_by_control_signal,
v5 - pump_off_by_power_off,
v6 - pump_off_by_control,
v7 - pump_blocked,
v8 - pump_on_by_control,

v9 - valve_open_signal_operating,
v10 - pump_on_signal_from_level_sensor_operating,
v11 - pump_on_signal_from_control_system_operating,
v12 - power_off,
v13 - valve_open_signal_from_level_sensor_on,

**<u>Input symptoms - elementary diagnoses:</u>**

d1 - valve_stuck_in_open_position_fault,
d2 -valve_open_control_signal_on,
d3 - level_sensor_on_when_level_too_high,
d4 - pump_on_by_control,
d5 - pump_broken_fault,
d6 - power_on.

### 4.4 Diagnostic Reasoning

In order to present the elements of diagnostic reasoning let us consider a particular diagnostic problem, e.g. one given by $M^{+} = \{m\}$ and $N^{+} = \{v2, d6\}$, i.e. the overflow of water when the system is switched on and the pump is observed not to work.

Before we start the analysis, let us propagate the observations through the graph. In fact, only d6=*true* can be propagated. We obtain v12=*false* and v5=*false*; this information can be used for checking consistency with hypothesised diagnoses.

With respect to our model there is one conjunctive cause possible, i.e. v1=*true* AND v2=*true*. The procedure must recursively explain both the symptoms found. Note that v2=*true* is consistent with the observations.

Now for explaining v1=*true* one can suggest two hypotheses, namely v3=*true* or v4=*true*. For explanation of v4=*true* again we have two possibilities; let assume that we select v13 to be *true* and thus we must have d3=*false* as the consequent selection; the process of explaining v1 is stopped here since we have arrived at an initial node being an elementary diagnosis. All the other possibilities of explaining v1=*true* are left for further possible use, while now the problem is to explain v2=*true*.

For v2=true there are three possibilities. Note that, however, the first one, i.e. v5=*true* is inconsistent with observations, so it should be left unexplored. Assume we select the second one, i.e. v6=*true*. Further on, the only explanation of v6/true is v8=*false*, which is consistent with the observations. And to explain v8=*false* we must assume that both v10=*false* and v11=*false* hold. This implies d3=*false* (already found on another way), and d4=*false*. The final diagnosis is then given by $D^{-} = \{d3, d4\}$, i.e. the level sensor is faulty and no control signal for setting on the pump is provided while the system is on. The further analysis of the diagnosis may lead to detection of one hard fault (element fault, d3=*false*), finding one control signal set to 0 (if this is a fault it should follow from a further analysis of the control algorithm for the specific conditions), and one operational condition already observed. The set of all possible diagnoses consistent with the observations found by the algorithm is:

({d1},{d3,d4}); ({d1},{d5}); ({},{d3,d4}); ({d5},{d3}); ({d2,d6},{d3,d4}); ({d2,d5,d6},{});

here any diagnosis is composed of two sets of elementary diagnoses: the *True* ones and the *False* ones. Note that only four of them are minimal diagnoses (i.e. 2,3,4 and 6) in the sense of pair wise set inclusion. But with respect to the sub-graph generated, all the six diagnoses constitute different solutions. For information, for the above problem specified with no initial observations there are as many as 14 potential solutions, and 6 of them are minimal with respect to set inclusion.

A sub-graph for the examined system is shown with slightly thicker lines in the figure below. The symptoms found true are marked with filled circles.

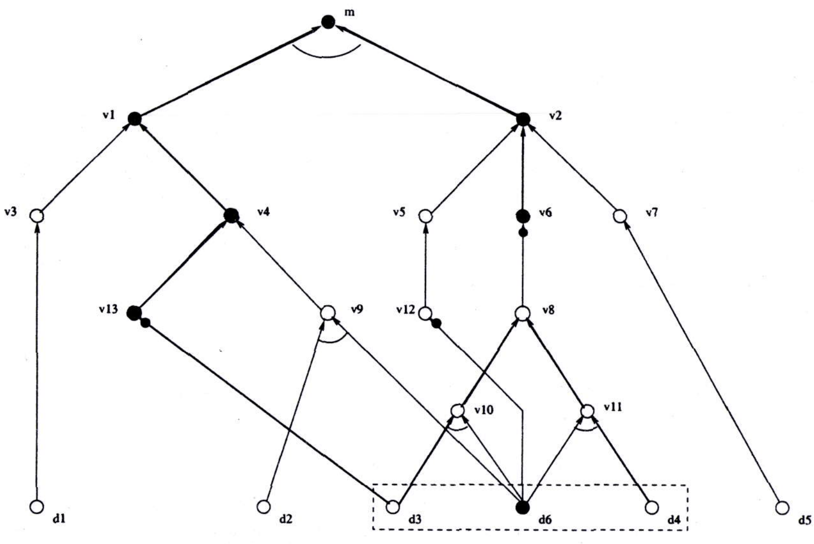


**Fig. 2:** A solution graph for the considered example diagnostic problem.

## 5 Concluding remarks

The main idea of the proposed approach is to put forward a relatively simple, straightforward formalism for diagnosis, easily understandable by domain experts, based on well studied techniques of AI, allowing for certain modifications and extensions. The proposed formalism constitutes a core, generic tools and can be extended and modified in a number of ways. Two extensions: qualitative probabilities and fuzzy degrees are pointed out.

The extension includes a tool that helps to graphically create and manipulate causal structures. The system was tested on a number of examples.

## References


Console L. and Torasso P. (1992). An approach to the compilation of operational knowledge from causal models. Edited by A. P. Sage. A publication of the IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-22, n.4, pp. 772-789.

De Kleer J. and Brown (1986). Theories of causal ordering. Artificial Intelligence. Editor-in-chief D. G. Bobrow. Published by Elsevier Science Publishers B. V., North-Holland, vol. 29, pp. 33-61.

Dubois D. and Prade H. (1985). Théorie des possibilités: applications à la représentation des connaissances en informatique. Ed. Masson.

Fuster-Parra P. and Ligeza A. (1995). Qualitative probabilities for causal diagnostic reasoning – Research an Development in Expert Systems. Presented at the 15th Annual Conference of the British Computer Society Specialist Group on Expert Systems. Edited by M.A. Bramer, J.L. Nealon and R. Milne. Published by SGES Publications, Information Press Ltd, Oxford, England, pp. 327-340, Cambridge, UK.

Fuster-Parra P. (1996). A model for causal diagnostic reasoning. Extended inference modes and efficiency problems. Ph. D. Thesis presented at the University of Balearic Islands.

Fuster-Parra P. and Ligeza A. (2000). Enhancing causality in an AND/OR/NOT causal graph for abductive diagnostic inference. Proceedings of the Fifteenth European Meeting on Cybernetics and Systems Research. Edited by Robert Trappl. Published by Austrian Society for Cibernetic Studies. Vienna, Austria, vol. 1, pp. 250-253.

Gaines, B.R. (1976). On the complexity of causal models. Edited by A. P. Sage. A publication of the IEEE Transactions on Systems, Man and Cybernetics, pp. 56-59.

Iwasaki Y. (1988). Causal ordering in a mixed structure. Proceedings of the 7th National Conference on Artificial Intelligence. Published by AAAI Press, The MIT Press, pp. 313-318.

Kuipers B. (1984). Commonsense reasoning about causality: deriving behavior from structure. Artificial Intelligence. Editor-in-chief D. G. Bobrow. Published by Elsevier Science Publishers B. V., North-Holland, vol. 24, pp. 169-203.

Nilsson N.L. (1971). Problem-Solving methods in artificial intelligence. McGraw-Hill Book Company, New York.

Pearl J. (1985). Heuristics. Intelligent Search strategies for computer problem solving. Addison-Wesley Publishing Company, Reading Massachusetts.

Pearl J. (1988). Embracing causality in default reasoning. Artificial Intelligence. Editor-in-chief D. G. Bobrow. Published by Elsevier Science Publishers B. V., North-Holland, vol.:35, pp. 259-271.

Zadeh L.A. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems. Published by Elsevier Science, Official Publication of the International Fuzzy Systems Association (IFSA), vol. 1, pp. 3-28.