Explanation of Bayesian networks and influence diagrams in Elvira

Carmen Lacave, Manuel Luque and Francisco Javier Díez

Abstract— **Bayesian networks and influence diagrams are probabilistic graphical models widely used for building diagnosis- and decision-support expert systems. Explanation of both the model and the reasoning is important for debugging these models, for alleviating users' reluctance to accept their advice, and for using them as tutoring systems. This paper describes some explanation options for Bayesian networks and influence diagrams that have been implemented in Elvira and how they have been used for building medical models and for teaching probabilistic reasoning to pre- and post-graduate students.**

Index Terms— **Bayesian networks, influence diagrams, expert systems, explanation, Elvira.**

I. INTRODUCTION

Bayesian networks (BNs) and influence diagrams (IDs) are two types of probabilistic graphical models widely used for building expert systems in several application domains. Both of them consist of acyclic directed graphs and probability distributions [1], [2], [3]. The main difference among them is that BNs only contain *chance nodes*, each representing a random variable, while IDs also contain *decision nodes*, which represent the options available to one or several decision makers, and *utility nodes*, which represent the decision makers' preferences. As a consequence, BNs can only be used in diagnostic problems, while IDs can be used as decision-support tools.

In the context of expert systems, either probabilistic or heuristic, the development of explanation facilities is important for three main reasons [4], [5]. First, because the construction of those systems with the help of human experts is a difficult and time-consuming task, prone to errors and omissions. An explanation tool can help the experts and the knowledge engineers taking part in the project to *debug* the system when it does not yield the expected results and even before a malfunction occurs. Second, because human beings are reluctant to *accept* the advice offered by a machine if they are not able to understand how the system arrived at those recommendations; this reluctancy is especially clear in medicine [6]. And third, because an expert system used as an intelligent *tutor* must be able to communicate the apprentice the knowledge it contains, the way in which the knowledge has been applied for arriving at a conclusion, and what would have happened if the user had introduced different pieces of evidence (what-if reasoning).

These reasons are especially relevant in the case of probabilistic expert systems, because the elicitation of probabilities is more difficult than the assessment of uncertainty in heuristic expert systems and because, even though probabilistic reasoning is just a formalization of (a part of) common-sense reasoning, the algorithms for the computation of probabilities and utilities are very different from the way a human being would draw conclusions from a probabilistic model.

Unfortunately, the explanation methods proposed so far are still unsatisfactory, as shown by the fact that most expert systems and commercial tools available today, either heuristic or probabilistic, have virtually no explanation capability [4], [7]. Despite the practical interest of this issue, very little research is currently carried out about explanation in probabilistic graphical models. As an attempt to palliate this shortcoming, in this paper we describe some methods for explaining both the model and the reasoning of probabilistic expert systems, which have been implemented in Elvira, a public software tool developed as a joint project of several Spanish universities. We also discuss how such methods respond to the needs that we have detected when building and debugging medical expert systems [8], [9], [10] and when teaching probabilistic graphical models to pre- and postgraduate students of computer science and medicine [11].

The rest of this paper is structured as follows: After reviewing the main features of explanation in expert systems in Section I-A and describing the Elvira software in Section I-B, we present the fundamentals of BNs and IDs in Sections II-A and II-B, respectively. Section III presents the facilities provided by Elvira for explaining both the model (Sec. III-A) and the reasoning (Sec. III-B) in BNs. Section IV analyzes how these facilities have been adapted for IDs in order to explain both the model (Sec. IV-A) and the results of inference (Sec. IV-B), to permit the introduction of evidence (Sec. IV-D), and to perform what-if reasoning with suboptimal policies (Sec. IV-E). The application of standard techniques, such as decision trees and sensitivity analysis, to explanation in IDs is discussed in Sections IV-C and IV-F, respectively. Section V analyzes related work and possible lines for future research, and Section VI presents the conclusions.

A. Features of explanation in expert systems

Explanation methods are characterized by several properties, corresponding to the main concepts on which an explanation is based [4], [7]: content, communication and adaptation. The

Manuscript received July 1, 2006; revised ...

Carmen Lacave is with the Dept. Technology and Information Systems, University of Castilla-La Mancha, Ciudad Real, 13071 Spain (e-mail: carmen.lacave@uclm.es). Manuel Luque and Francisco Javier Díez are with the Dept. Artificial Intelligence, UNED, Madrid, 28040 Spain (e-mail: mluque@bec.uned.es, fjdiez@dia.uned.es). They are all members of the Research Center on Intelligent Decision-Support Systems (CISIAD), at UNED.

content of an explanation deals with either the model, the reasoning, or the available evidence. Explanation of the model, also known as *static explanation* [12], consists in showing the information represented by the knowledge base of the expert system in a way that it can be easily understood by the user. Explanation of the reasoning, or *dynamic explanation*, describes how and why the system has obtained certain results. Explanation of evidence usually consists in finding the most probable configuration that justify the evidence [1], which is also known as *abduction*. Dynamic explanations can be generated at the micro or the macro level [13]: micro-level explanations try to justify why the probability of a certain variable has varied, why the belief on a certain hypothesis has changed, or why a rule has fired as a consequence of the variations in its neighbor variables or rules; on the contrary, macro-level explanations analyze the main lines of reasoning (the paths in the Bayesian network, the chains of rules, etc.) that led from the evidence to a certain conclusion.

The second main aspect of explanation, namely communication, is related to the way of interacting with the user and the way of presenting the explanations, either textually or graphically or by a combination of both.

Finally, adaptation refers to the ability to modify the explanations and the interaction depending on the user's expertise and needs. See [4], [7] for a more detailed analysis of these features and for a detailed review of the most relevant methods and systems offering some kind for explanation, both for Bayesian networks [4] and for heuristic expert systems [7].

B. Elvira

Elvira[1] is a tool for building and evaluating graphical probabilistic models [14]. It resulted from a joint research project of several Spanish universities. It is implemented in Java, so that it can run on different platforms. It contains a graphical interface for editing networks, with specific options for canonical models (e.g., OR, AND, MAX...), exact and approximate algorithms for discrete and continuous variables, explanation facilities, learning methods for building networks from databases, algorithms for fusing networks, etc. Although some of the algorithms work with both discrete and continuous variables, the explanation capabilities assume that all the variables are discrete.

a) Architecture of Elvira: Elvira is structured in four main modules:

- Data representation, which contains the definition of the data structures needed for managing BNs and IDs in Java.
- Data acquisition, including the classes necessary for saving and loading a network both from a file and from a data base, the parser, etc. It also contains classes for exporting and importing the networks in several formats.
- Processing. This module implements the algorithms for processing and evaluating the models. It is organized in several submodules, one for each task: inference, learning, fusion, decision trees, sensitivity analysis...
- Visualization, which mainly defines the Elvira GUI and, obviously, makes use of the classes included in the previous modules. This module contains the classes for generating explanations and for the *internationalization* of the whole program, i.e., for displaying the interface in different languages. Currently, only Spanish and English are supported, but other languages can be easily added.

The main advantages of this modular design is that each group involved in the project can focus on a different task and that the program can be easily extended with new functionalities or adapted to different needs.

b) Working with the Elvira GUI: In addition to invoking Elvira's classes from the command line and using it as an API, it is possible to interact with Elvira by means of its GUI, which has two working modes:

- edit, for graphically editing BNs and IDs. This is possible by means of several windows which help the user to build or to modify the model manually, by requesting all the data associated to the nodes, the arcs and the properties of the whole BN or ID. Alternatively, BNs can be built from data bases by applying some of the many learning algorithms implemented in Elvira; and
- inference, for propagating evidence and explaining the results. The introduction of evidence can be done by clicking on the node, as in other software tools, or by means of an editor of cases [15], which provides a list of the variables in the model. With respect to the inference process, the user can choose one of several algorithms, with many variations, and in the case of a BN, they can select either evidence propagation or abduction[2] and whether the evidence is propagated automatically (i.e., just after the user introduces or removes a finding) or manually (by demand). Most of the explanation capabilities provided by Elvira (see below) are offered in the inference mode.

II. MATHEMATICAL FOUNDATIONS

Before giving the definition of Bayesian network (BN) and influence diagram (ID), we establish some notational conventions.

Notation: Given that each node in a probabilistic graphical model represents a variable, in this paper we will use both terms indifferently. We will use a capital letter V to represent a variable, and its corresponding lower case letter v for representing a generic value. Sets of variables will be represented by bold capital letters $\mathbf{V}$, and a bold lower case letter $\mathbf{v}$ will represent a configuration of values of $\mathbf{V}$. In the context of directed graphs, $Pa(V)$ represents the set of parents of node V, and $pa(V)$ a configuration of the variables in $Pa(V)$.

Definitions: A *finding f* is a piece of information that states with certainty the value taken on by a chance variable. A

[1] At `http://www.ia.uned.es/~elvira` it is possible to obtain the source code and several technical documents about Elvira.

[2] In the context of BNs, *evidence propagation* usually refers to computing the posterior probability of each single variable given the available evidence, while *abduction* consists in computing the joint probability of a set of variables of interest given the evidence, what is also called *Maximum A Posteriori Probability (MAP).*

finding may be, for example, the fact that the patient is a male; other findings might be that he is 54 years old, he has fever, he does not usually have headaches, etc. The set of findings is called *evidence* and corresponds to a certain configuration $\mathbf{e}$ of the observed variables $\mathbf{E}$.

A. Bayesian networks

A BN consists of an acyclic directed graph (ADG), whose nodes represent a set $\mathbf{V}_C$ of chance variables, where C stands for "chance", and whose links represent —roughly speaking— probabilistic dependencies among them, together with a probability distribution over its variables that satisfies the *d*-separation property [1]. This property implies that the joint probability distribution can be factored as the product of the probability of each node conditioned on its parents.

$$P(\mathbf{v}_C) = \prod_{V \in \mathbf{V}_C} P(v|pa(V)) \tag{1}$$

As a consequence, the quantitative information of a Bayesian network can be given by assigning to each chance node C a probability distribution $P(c|pa(C))$ for each configuration of its parents, $pa(C)$. Both the graph and the probabilities of a BN can be obtained *automatically*, from data bases, or *manually*, from human experts' knowledge and the literature for the domain to be modeled. In this case, the elicitation of probabilities constitutes a very difficult task, usually referred to as a bottleneck in the development of BNs [16].

Probabilistic reasoning in BNs usually consists in computing the posterior probability of some variables of interest $\mathbf{V}_I \subseteq \mathbf{V}_C \setminus \mathbf{E}$ given the available evidence, $P(\mathbf{v}_I|\mathbf{e})$.

B. Influence Diagrams

1) Definition of an ID: An influence diagram (ID) contains three kinds of nodes: *chance nodes* $\mathbf{V}_C$, *decision nodes* $\mathbf{V}_D$, and *utility nodes* $\mathbf{V}_U$—see Fig. 1. Chance nodes represent events not controlled by the decision maker. Decision nodes correspond to actions under the direct control of the decision maker. Utility nodes represent the decision maker's preferences. Utility nodes can not be parents of chance or decision nodes.

In the extended framework proposed by Tatman and Shachter [17] there are two kinds of utility nodes: *ordinary utility nodes*, whose parents are decision and/or chance nodes (such as U_1 and U_2 in Fig. 1), and *super-value nodes*, whose parents are utility nodes (U_0 in Fig. 1 is a super-value node). We assume that there is a utility node that is either the only utility node or a descendant of all the other utility nodes, and therefore has no children; we denote it by U_0.[3]

There are three kinds of arcs in an ID, depending on the type of node they go into. Arcs into chance nodes represent probabilistic dependency. Arcs into decision nodes represent availability of information, i.e., an arc $Y \rightarrow D$ means that the state of Y is known when making decision D. Arcs into utility nodes represent functional dependence: for ordinary utility nodes, they represent the domain of the associated utility function; for a super-value node they indicate that the associated utility is a function (usually the sum or the product) of the utility functions of its parents.

Standard IDs require that there is a directed path that includes all the decision nodes and indicates the order in which the decisions are made. This in turn induces a partition of $\mathbf{V}_C$ such that for an ID having n decisions $\{D_0, \ldots, D_{n-1}\}$, the partition contains $n+1$ subsets $\{\mathbf{C}_0, \mathbf{C}_1, ..., \mathbf{C}_n\}$, where $\mathbf{C}_i$ is the set of chance variables C such that there is a link $C \rightarrow D_i$ and no link $C \rightarrow D_j$ with $j < i$; i.e., $\mathbf{C}_i$ represents the set of chance variables known for D_i and unknown for previous decisions. $\mathbf{C}_n$ is the set of variables having no link to any decision, i.e., the variables whose true value is never known directly. In our example (Fig. 1), $D_0 = T$, $D_1 = D$, $\mathbf{C}_0 = \varnothing$, $\mathbf{C}_1 = \{Y\}$, and $\mathbf{C}_2 = \{X\}$.

The variables known to the decision maker when deciding on D_i are called *informational predecessors* of D_i and denoted by *IPred*(D_i). Standard IDs assume the *no-forgetting hypothesis*, which means that the decision maker remembers all previous observations and decisions. By assuming such property we have

$$\begin{aligned} \textit{IPred}(D_i) &= \textit{IPred}(D_{i-1}) \cup \{D_{i-1}\} \cup \mathbf{C}_i & (2)\\ &= \mathbf{C}_0 \cup \{D_0\} \cup \mathbf{C}_1 \cup \ldots \cup \{D_{i-1}\} \cup \mathbf{C}_i \,. & (3) \end{aligned}$$

An arc $V \rightarrow D$, where D is a decision and V is either a decision or a chance node, is said to be *non-forgetting* is there is another directed path from V to D. In standard IDs non-forgetting arcs are irrelevant: they can be added or removed without changing the semantics of the ID.

The quantitative information that defines an ID is given by assigning to each chance node C a probability distribution $P(c|pa(C))$ for each configuration of its parents (as in the case of BNs), assigning to each ordinary utility node U a function $\psi_U(pa(U))$ that maps each configuration of its parents onto a real number, and assigning a utility-combination function to each super-value node. The domain of each function U is given by its *functional predecessors*, *FPred*(U). For an ordinary utility node, *FPred*$(U) = Pa(U)$, and for a super-value node *FPred*$(U) = \bigcup_{U' \in Pa(U)}$*FPred*$(U')$. In the above example, *FPred*$(U_1) = \{X, D\}$, *FPred*$(U_2) = \{T\}$, and *FPred*$(U_0) = \{X, D, T\}$. In order to simplify the notation, we assume without loss of generality that *FPred*$(U_0) = \mathbf{V}_C \cup \mathbf{V}_D$.

For each configuration $\mathbf{v}_D$ of the decision variables $\mathbf{V}_D$ we have a joint distribution over the set of chance variables $\mathbf{V}_C$:

$$P(\mathbf{v}_C : \mathbf{v}_D) = \prod_{C \in \mathbf{V}_C} P(c|pa(C)) \tag{4}$$

which represents the probability of configuration $\mathbf{v}_C$ when the decision variables are externally set to the values given by $\mathbf{v}_D$ [18].

2) Policies and expected utilities: A *stochastic policy* for a decision D is a probability distribution defined over D and conditioned on the set of its informational predecessors,

[3] An ID that does not fulfill this condition can be transformed by adding a super-value node U_0 of type sum whose parents are the utility nodes that did not have descendants. The expected utility and the optimal strategy of the transformed diagram are the same as those of the original one.

$P_D(d|iPred(D))$. If P_D is degenerate (consisting of ones and zeros only) then we say that the policy is deterministic.

A *strategy* Δ for an ID is a set of policies, one for each decision, $\{P_D | D \in \mathbf{V}_D\}$. A strategy Δ *induces* a joint distribution over $\mathbf{V}_C \cup \mathbf{V}_D$ defined by

$$\begin{aligned} &P_\Delta(\mathbf{v}_C, \mathbf{v}_D) \\ &\quad = P(\mathbf{v}_C : \mathbf{v}_D) \prod_{D \in \mathbf{V}_D} P_D(d|\mathit{IPred}(D)) \\ &\quad = \prod_{C \in \mathbf{V}_C} P(c|pa(C)) \prod_{D \in \mathbf{V}_D} P_D(d|pa(D)) \end{aligned} \tag{5}$$

Let I be an ID, Δ a strategy for I and $\mathbf{r}$ a configuration defined over a set of variables $\mathbf{R} \subseteq \mathbf{V}_C \cup \mathbf{V}_D$ such that $P_\Delta(\mathbf{r}) \neq 0$. The conditional probability distribution *induced by strategy* Δ *given the configuration* $\mathbf{r}$, defined over $\mathbf{R}' = (\mathbf{V}_C \cup \mathbf{V}_D) \setminus \mathbf{R}$, is given by:

$$P_\Delta(\mathbf{r}'|\mathbf{r}) = \frac{P_\Delta(\mathbf{r}, \mathbf{r}')}{P_\Delta(\mathbf{r})} \tag{6}$$

Using this distribution we can compute the *expected utility of* U *under strategy* Δ *given the configuration* $\mathbf{r}$ as:

$$EU_U(\Delta, \mathbf{r}) = \sum_{\mathbf{r}'} P_\Delta(\mathbf{r}'|\mathbf{r}) \psi_U(\mathbf{r}, \mathbf{r}') \tag{7}$$

For the terminal utility node U_0, $EU_{U_0}(\Delta, \mathbf{r})$ is said to be the *expected utility of strategy* Δ *given the configuration* $\mathbf{r}$, and denoted by $EU(\Delta, \mathbf{r})$.

We define the *expected utility of* U *under strategy* Δ as $EU_U(\Delta) = EU_U(\Delta, \blacklozenge)$, where $\blacklozenge$ is the empty configuration. We have that

$$EU_U(\Delta) = \sum_{\mathbf{v}_C} \sum_{\mathbf{v}_D} P(\mathbf{v}_C, \mathbf{v}_D) \psi_U(\mathbf{v}_C, \mathbf{v}_D) \tag{8}$$

We also define the *expected utility of strategy* Δ as $EU(\Delta) = EU_{U_0}(\Delta)$.

An *optimal strategy* is a strategy Δ_{opt} that maximizes the expected utility:

$$\Delta_{opt} = \arg\max_{\Delta \in \Delta^*} EU(\Delta) \tag{9}$$

where Δ^* is the set of all strategies for I. Each policy in an optimal strategy is said to be an *optimal policy*. The *maximum expected utility* (*MEU*) is

$$\mathit{MEU} = EU(\Delta_{opt}) = \max_{\Delta \in \Delta^*} EU(\Delta) \tag{10}$$

The evaluation of an ID consists in finding the *MEU* and an optimal strategy, composed by an optimal policy for each decision. It can be proved [18], [3] that

$$\mathit{MEU} = \sum_{\mathbf{c}_0} \max_{d_0} \ldots \sum_{\mathbf{c}_{n-1}} \max_{d_{n-1}} \sum_{\mathbf{c}_n} P(\mathbf{v}_C : \mathbf{v}_D) \psi(\mathbf{v}_C, \mathbf{v}_D) \tag{11}$$

For instance, the *MEU* for the ID in Fig. 1 is

$$\begin{aligned} \mathit{MEU} = &\max_t \sum_y \max_d \sum_x \\ &P(x) \cdot P(y|t, x) \cdot \underbrace{(U_1(x, d) + U_2(t))}_{U_0(x,d,t)} \end{aligned} \tag{12}$$

3) Cooper policy networks: A strategy $\Delta = \{P_D | D \in \mathbf{V}_D\}$ can be used to convert the ID into a BN, that we call *Cooper policy network* (CPN), as follows: each decision D is replaced by a chance node with probability potential P_D and parents $\mathit{IPred}(D)$, and each utility node U is converted into a chance node whose parents are its functional predecessors, $\mathit{FPred}(U)$—see Fig. 2. The values of each new chance variable U are $\{+u, \neg u\}$ and its probability is $P_{\mathit{CPN}}(+u|\mathit{fPred}(U)) = norm_U(U(\mathit{fPred}(U)))$, where $norm_U$ is a linear transformation that maps the utilities $U(\mathit{fPred}(U))$ from the interval $[\alpha_U, \beta_U]$ onto the interval $[0, 1]$ [19]; α_U and β_U are defined as:

$$\alpha_U = \min_{\mathit{fPred}(U)} \psi_U(\mathit{fPred}(U)) \tag{13}$$

$$\beta_U = \max_{\mathit{fPred}(U)} \psi_U(\mathit{fPred}(U)) \; . \tag{14}$$

The joint distribution of the CPN is:

$$\begin{aligned} &P_{\mathit{CPN}}(\mathbf{v}_C, \mathbf{v}_D, \mathbf{v}_U) \\ &\qquad = P_\Delta(\mathbf{v}_C, \mathbf{v}_D) \prod_{U \in \mathbf{V}_U} P_U(u|pa(U)) \end{aligned} \tag{15}$$

Given two configurations $\mathbf{r}$ and $\mathbf{r}'$defined over two set of variables, $\mathbf{R} \subseteq \mathbf{V}_C \cup \mathbf{V}_D$ and $\mathbf{R}' \subseteq (\mathbf{V}_C \cup \mathbf{V}_D)$, such that $\mathbf{R} \cap \mathbf{R}' = \varnothing$ and $P(\mathbf{r}) \neq 0$, and U a utility node, it holds that

$$P_\Delta(\mathbf{r}') = P_{\mathit{CPN}}(\mathbf{r}') \tag{16}$$

$$P_\Delta(\mathbf{r}'|\mathbf{r}) = P_{\mathit{CPN}}(\mathbf{r}'|\mathbf{r}) \tag{17}$$

$$EU_U(\Delta) = norm_U^{-1}(P_{\mathit{CPN}}(+u)) \tag{18}$$

$$EU_U(\Delta, \mathbf{r}) = norm_U^{-1}(P_{\mathit{CPN}}(+u|\mathbf{r})) \tag{19}$$

In Section IV we will use these equations to compute on a CPN the probabilities and expected utilities to be displayed in the GUI.

III. EXPLANATION OF BAYESIAN NETWORKS IN ELVIRA

This section describes the main options available in Elvira for generating explanations of both the model and the reasoning.

A. Explanation of the model

Elvira offers verbal and graphical explanations at the micro level of given nodes and links (cf. Sec. I-A), and also of the whole network, by means of windows and menus, as follows. Currently Elvira treats all variables as if they were ordinal. In the case of non-ordinal variables, such as sex or race, the order is that in the list of states defined by the user while editing the network.

1) Explanation of nodes: In edit mode, nodes are displayed as *contracted* [15], i.e., drawn as an oval containing only its name. However, in inference mode, nodes can also be displayed as *expanded*, i.e., drawn as rounded-corner rectangles which graphically display the main properties of the nodes (states and its probabilities). For example, in the BN in Fig. 3 the nodes Virus A, Virus B, Disease 1 and Disease 2 are expanded and the rest are contracted.

The verbal explanation of a given node, which can be accessed by right-clicking on it, contains the following information: name, states, parents and children, prior odds and posterior odds. This is very useful for analyzing the correctness of some probabilities, since in some cases human experts know that a value is certain times more probable than other, instead of the concrete data. Also the verbal explanation of a node includes some other properties, such as the purpose and the importance factor [15] of such node. The *purpose* of a node is defined by the role it plays in the model, according to several categories, such as "symptom" or "disease". The *importance factor*, a value assigned by the human expert on a 0–10 scale, is the same as the relevance factor used in DIAVAL [8] for selecting the main diagnoses and equivalent to the importance factor in MYCIN [20]. Additionally, the importance factor can work in conjunction with the *expansion threshold* set by the user [5]—in Fig. 3 it is set to 7.00 (see the upper left corner of the figure). The nodes whose importance factor is higher than the expansion threshold and whose role is one of those selected by the user are expanded, and the rest are contracted. In Fig. 3 the only selected role (determined by means of a specific screen) was disease. It is also possible to manually expand or contract a particular node.

The facility of selectively expanding nodes has been very useful when building and debugging real-world models containing a high number of nodes, such as Prostanet, a BN for diagnosing prostate cancer [5], and Hepar II, a BN for the diagnosis of liver disorders [21]. For example, when evaluating Prostanet, which contains 47 nodes, the expert wanted to focus only on the probabilities of the main diseases, in order to make a differential diagnosis between prostate cancer and some other benign diseases related to prostate. We could do it by automatically expanding the nodes whose purpose was disease/anomaly and whose importance factor was greater than 7.

In a similar way to DIAVAL [8], Elvira allows the user to navigate across the explanation windows associated to the nodes and links of the network in order to analyze at a micro level all the information related to each of them. This facility is not necessary for networks containing "only" a few dozens nodes, because the graph can be seen on a screen, but may be useful for bigger networks with intricate graphs.

2) Explanation of links: One of the more useful features of Elvira is the automatic coloring of links [5], which offers qualitative insight about the conditional probability tables. This coloring is based on the sign of influence [22] and the magnitude of influence [5], which are defined as follows:

Definition 1: Let A and C be two ordinal variables such that the former is one of the parents of the latter, $Pa(C) = \{A\} \cup \mathbf{B}$. The *magnitude of the influence* (MI) for link $A \rightarrow C$ is

$$MI(A, C) = \max_{c,a,\mathbf{b}} |P(C \geq c|a, \mathbf{b}) - P(C \geq c|a_0, \mathbf{b})| \quad (20)$$

where a_0 is the *normal* value of A.

The normal value of a variable is the state that represents the absence of anomaly. For instance, if X represents a disease having a domain {present, absent} or {severe, moderate, mild, absent}, the normal value is "absent". If the domain is {increased, normal, decreased}, the normal value is "normal". Therefore, the $MI(A, C)$ measures to what extent a certain cause A is able to shift C from its normality state to a state of anomaly.

Definition 2: If A and C are ordinal variables and $Pa(C) = \{A\} \cup \mathbf{B}$, we say that A *positively influences* variable C iff $MI(A, C) \neq 0$ and

$$\begin{aligned} \forall c, \forall a, \forall a', \forall \mathbf{b}, \quad a > a' \Longrightarrow \\ P(C \geq c|a, \mathbf{b}) \geq P(C \geq c|a', \mathbf{b}) \end{aligned} \quad (21)$$

We also say that the link is *positive*.

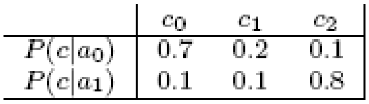
The intuition motivating these definitions is that an influence is positive when higher values of A make high values of C more probable, as shown in the following example.

Example 3: If A and C are binary, the ordering $+a > \neg a$ and $+c > \neg c$ implies that $P(C \geq +c|a, \mathbf{b}) = P(+c|a, \mathbf{b})$ and $P(C \geq \neg c|a, \mathbf{b}) = P(+c|a, \mathbf{b}) + P(\neg c|a, \mathbf{b}) = 1$. Therefore $P(C \geq \neg c|a, \mathbf{b}) = P(C \geq \neg c|a', \mathbf{b})$ in all cases. Consequently, A positively influences C iff $P(+c|+a, \mathbf{b}) \geq P(+c|\neg a, \mathbf{b})$ for all $\mathbf{b}$'s and the inequality holds strictly for at least one $\mathbf{b}$.

The reason for using $P(C \geq c|a, \mathbf{b})$ instead of $P(c|a, \mathbf{b})$ in the above definitions becomes clear by observing the example in Table I, in which a_1 clearly leads to higher values of C than a_0; it is a case of positive influence according with Definition 2, but it would have not been so if we had used "$P(c|a, \mathbf{b}) \geq P(c|a', \mathbf{b})$" in the definition, because $P(c_0|a_1) < P(c_0|a_0)$ and $P(c_1|a_1) < P(c_1|a_0)$.

TABLE I

A PROBABILITY TABLE SHOWING A POSITIVE INFLUENCE OF A ON C (SEE DEF. 2, EQ. 21).



	c_0	c_1	c_2
$P(c\|a_0)$	0.7	0.2	0.1
$P(c\|a_1)$	0.1	0.1	0.8

The definitions of *negative influence* and *negative link* are analogous. When $MI(A, C) = 0$ we say that the influence of link $A \rightarrow C$ is *null*. From a point of view of knowledge representation, a BN should not contain null links. When the influence is neither positive nor negative nor null, then it is said to be *undefined*. A link $A \rightarrow C$ may be undefined for several reasons. One of them is the case in which A has more than two values and the cumulative probability of C increases when A changes from a_0 to a_1 but decreases when changing from a_1 to a_2. For instance, the probability of prostate cancer increases until a man is in his 50's and decreases afterwards. A link can also be undefined if A increases the probability of high values of C for some configurations of $\mathbf{B}$ and decreases it for other configurations. For instance, a certain drug might be beneficial for a type of patients and harmful for others.

Positive links are colored in red, negative in blue, undefined in purple, and null in black.[4] In Fig. 3 we can see that most links are red, because in general the presence of the cause increases the probability of the effect; the only exception in that example is the link Vaccination⟶Disease2, for obvious reasons. We can also see in that figure that the thickness of links varies with the magnitude of influence, i.e., with the strength of the association.

The coloring and the width of links is one of the most powerful tools provided by Elvira in order to help both experts and knowledge engineers to detect wrong influences, which frequently occurs when probabilities are subjectively estimated, and even when the probabilities are obtained from databases, either because of several biases or because of values missing non-randomly. In fact, without Elvira's graphical explanation of the links it would have been much more difficult to detected some wrong probabilities estimated by the experts when building Prostanet [5] and Hepar II [21]. Elvira also allowed us to see at a glance that many of the influences in the database version of Hepar II (i.e., the model in which all the probabilities were drawn from a database) were negative [21], [23], which seriously questioned the validity of that database as a source of information for building BNs.[5]

Additionally, Elvira can offer verbal explanations for a selected link, which we do not describe here because of the lack of space. The interested reader is referred to [15].

3) Explanation of the network: Elvira can generate a verbal explanation of the whole network based mainly on the purpose of each node. It consists of a text containing a description of the disease/anomaly nodes, based on their parents and children. For example, a fragment of the verbal explanation of the network in Fig. 3 is: *The network "Two Diseases" represents the following information: The disease / anomaly Virus A has neither causes nor risk factors represented in the network. It may cause the following DISEASES / ANOMALIES: Disease 1, SYMPTOMS: Symptom, SIGNS: Sign.* This tool has been very useful when building the causal graph of Prostanet [5], because the natural-language texts provided by Elvira, which is similar to the way in which the network would be described by a human being, helped the experts to understand the causal model represented by the graph of the BN.

B. Explanation of reasoning: evidence cases

1) Explanation of an evidence case: In Elvira an evidence case is defined as a set of findings plus the corresponding posterior probabilities:

Definition 4: Given a Bayesian network, defined over a set of variables $\mathbf{V}_C$**,** and evidence $\mathbf{e}$, an *evidence case* (EC) is a pair $(\mathbf{e}, \mathbf{P}^*)$, where $\mathbf{e}$ is the configuration of the observed variables $\mathbf{E}$ that represents the set of findings, and $\mathbf{P}^*$ is the set of posterior probabilities of the unobserved nodes: $\mathbf{P}^* = \{P(V|\mathbf{e}), V \in \mathbf{V_C} \setminus \mathbf{E}\}$.

The individual probabilities $P(v|\mathbf{e})$ can be observed by inspecting the posterior probability of each value in the expanded nodes, as shown in Fig. 3. Additionally, it is possible to have an overall view of the changes in the probabilities all the variables by selecting *automatic explanation* in the explanation options menu, which performs a coloring of nodes depending on the changes of their posterior probabilities, in accordance with the following definitions.

Definition 5: Evidence $\mathbf{e}$ **influences** variable V more than θ (with $\theta \geq 0$) iff

$$\exists v, |P(V \geq v|\mathbf{e}) - P(V \geq v)| > \theta, \tag{22}$$

When the influence exists, it is said to be **positive** iff

$$\forall v, P(V \geq v|\mathbf{e}) \geq P(V \geq v) \tag{23}$$

and it is **negative** iff

$$\forall v, P(V \geq v|\mathbf{e}) \leq P(V \geq v) \tag{24}$$

If $\mathbf{e}$ does not influence V, we can say that the influence is null. A non-null influence that is neither positive nor negative is **undefined**. In the case of a binary variable V, the influence is positive if $P(+v|\mathbf{e}) > P(+v)$, negative if $P(+v|\mathbf{e}) < P(+v)$ and null if $P(+v|\mathbf{e}) = P(+v)$.

This definition, as well as those for the coloring of links in Sec. III-B.2, is based on Wellman's work on qualitative probabilistic networks, *QPN* [22]. However, the fact that our networks contain numerical probabilities and that propagation of evidence is done by quantitative algorithms, allows us to determine the sign of probability changes in many cases in which Wellman's algorithms would lead to "unknown" signs.

The quantitative aspect of influence is measured by the following magnitude:

Definition 6: If evidence $\mathbf{e}$ influences variable V, we define the **magnitude of the impact of evidence** $\mathbf{e}$ **over** V, as

$$MI_{\mathbf{e}}(V) = \max_v |P(V \geq v|\mathbf{e}) - P(V \geq v)| \tag{25}$$

If V is a binary variable then $MI_{\mathbf{e}}(V) = |P(+v|\mathbf{e}) - P(+v)| = |P(\neg v|\mathbf{e}) - P(\neg v)|$.

In Elvira, nodes are colored in red if they receive positive influence from $\mathbf{e}$, in blue if the influence is negative, and in

[4] The coding of influences and probabilities in Elvira is inspired in physics, where high temperatures are associated with the red color and low temperatures with blue.

[5] Because of our experience in the field of medical applications, we suspect that the quality of some of the databases used for building BNs with learning algorithms may suffer from similar biases. In that case, a cross-validation of the model (against another portion of the database) does not at all mean that the resulting BN represents the real-world correlations and influences.

purple if it is undefined. If the influence is null, they remain colored in yellow, the default color. The saturation of the color depends on the magnitude of the impact of evidence. The threshold θ (cf. Eq. 22) can be set from the GUI.

The coloring of nodes is especially useful to analyze the propagation of evidence along different chains of reasoning [24]. For instance, in Fig. 4 we can see how the finding X-ray=positive propagates up to variable Vaccination and why it causes a decrease of $P(\text{vaccination})$. The coloring of nodes and links offers an intuitive idea of how the probabilities have changed due to evidence propagation. The fact that link Anomaly→X-ray is positive explains why a positive finding for X-ray leads to an increase in the probability of Anomaly, which is colored in red—see the rules for the combination of signs in [22]. The same explanation applies to the positive link Disease 2→Anomaly. On the contrary, the link Vaccination→Disease 2 (depicted in blue) is negative, and this explains why an increase in the probability of Disease 2 makes us suspect that the patient was not vaccinated, which is reflected in the blue coloring of node Vaccination.

Additionally, Elvira is able to classify the findings depending on the kind and magnitude of influence that they exert on a certain variable V, selected by the user, according to the following definitions:

Definition 7: Given evidence $\mathbf{e}$, the **magnitude of influence** exerted by a finding f over variable V is

$$MI_f(V) = \max_v |P(V \geq v|\mathbf{e}) - P(V \geq v|\mathbf{e} \setminus \{f\})| \quad (26)$$

In this context, we say a finding f *positively influences* variable V iff $MI_f(V) \neq 0$ and

$$\forall v, P(V \geq v|\mathbf{e}) > P(V \geq v|\mathbf{e} \setminus \{f\}) \quad (27)$$

The definitions of *negative* and *null* influence are similar. When a non-null influence is neither positive nor negative, then it is said to be *undefined*. Please note that Definitions 5 and 6 refer to the impact of a set of evidence as a whole, while Definition 7 focuses on the impact of an individual finding (in the context of other findings).

The classification of findings allows the user to understand why the probability of a node receiving different kinds of influence has increased or decreased. For example, given the network in Fig. 3, if the user introduces the evidence {Vaccination=yes, Anomaly=present}, the probability of Disease 2 increases, as shown by the red coloring of this node. The classification of findings helps the user to understand that the positive influence of Vaccination=yes, whose magnitude is 0.493, prevails over the negative influence of Anomaly=present, whose magnitude is only 0.145, as it is shown in Fig. 5.

These explanation options can be accessed by opening a window which contains the following information about the current evidence case:

1) The *name* of the case and its associated *findings*.
2) The *probability of evidence* $\mathbf{e}$. This may be useful, for instance in medicine, because diagnosing a rare disease can be explained by a low value of $P(\mathbf{e})$. It can also be used to detect conflicts between findings [25].
3) A panel for the *analysis of sensitivity to the evidence*.[6] This panel shows the states of a certain variable V selected by the user, their prior probabilities, their posterior probabilities and the logarithmic ratio of both probabilities for each state v:

$$S(v|\mathbf{e}) = \lg \left(\frac{P(v|\mathbf{e})}{P(v)} \right) \quad (28)$$

We have chosen this function because its meaning can be easily explained to users: positive values mean that the probability of v has increased, and vice versa, and the absolute value of S measures the impact of the evidence on v.
4) Two buttons, How and Why, whose names are inspired in MYCIN's explanation facilities [31]. The How button highlights the chains of reasoning by hiding the links and nodes that are in no active path from the evidence nodes $\mathbf{E}$ to variable V, and colors the nodes in active paths, as explained in the previous section. The decision of whether a path is active, inactive, or blocked is based on the d-separation criteria [1]. In turn, the Why button opens a window having four list of findings, depending on whether the influence exerted by each one on V is positive, negative, null, or undefined, according with Equation 7, as we have described earlier and illustrated by Fig. 5.

Then, the coloring of the nodes in the paths from the findings to a given variable V helps the user to analyze how evidence flows through the network, increasing or decreasing the probability of some variables in the way up to V, with different degrees of intensity. The classification of findings also helps the users to detect the findings that have more impact than others and also the possible conflicts among findings. The study of different evidence cases (see below) allows the user to analyze the impact of each finding by itself and its impact in the context of a whole set of evidence.

2) Handling several evidence cases: One of the specific features of Elvira, which differentiates it from the tools developed previously, is its ability to manage several evidence cases simultaneously [15]. By default, Elvira creates a *prior case*, which corresponds to the absence of evidence, and whose associated probabilities are the prior probabilities of each node in the network: $(\varnothing, \{P(V)|V \in \mathbf{V_C}\})$. It also generates a new case automatically when the user introduces the first finding. Later, the user can create new cases to accommodate new sets of findings; the new case inherits all the findings of the previous one.

Most of the operations are performed on the *current* case, which can be selected by the user among the list of available

[6] There are two kinds of sensitivity analysis in BNs. *Sensitivity to the evidence*, which we are discussing now, refers to how the set of findings has affected the posterior probabilities [24], [26]. In contrast, the analysis of *sensitivity to the parameters* studies how different variations of the conditional probabilities that define the network would affect the prior and posterior probabilities [27], [28], [29], [30]. In Section IV-F we will discuss the analysis of sensitivity to the parameters in IDs.

cases. A different bar is displayed for each evidence case and each state in the expanded nodes, although the numerical probability is displayed only for the current case. In Fig. 3 four evidence cases are considered—this is why each expanded node has four colored bars associated to each state. The first one is colored in green and it represents the prior probabilities of each node, i.e, the absence of evidence. Nodes whose value is known with certainty, because the current evidence case contains one finding associated to it, are colored in gray. In Fig. 3 the there is only one gray node, Symptom, representing the only finding of the current case. Also the tool bar shows a label with the same color as the current case and its name. For example, in Fig. 3 the current case corresponds to the one colored in red and identified as "Presence of Symptom". If every node were expanded the user could identify the findings associated to the rest of cases because the corresponding colored bars were set to the maximum possible length.

A *monitor of cases* permits the user to control which cases are stored and displayed, to remove cases from memory, and to modify some properties of the cases, such as the name and color that Elvira assigns by default, which helps the user to easily identify each evidence case. For example, in Fig. 3 the current case has been renamed as "Presence of symptom". An *editor of cases* makes it possible to navigate through the different evidence cases stored in the memory, to edit each of them, and to propagate new evidence.

The analysis of the propagation of evidence through chains of reasoning, combined with the possibility of handling several evidence cases simultaneously, has been very useful for our students to have an intuitive understanding of the d-separation properties, a notion that was quite complicated for them before we used Elvira in our tuition. Using some example networks, we illustrate, for instance, the difference between active and inactive paths by showing how the introduction of a certain finding changes the color of some nodes, while others remain in yellow. This change or lack of change in the probabilities can also be seen (when the nodes are expanded) by observing the probability bars. Similarly, the fact that two sets of variables, $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ can be illustrated by first introducing a finding for each variable in $\mathbf{Z}$ and then creating a new case that, in addition, contains evidence for some of the variables in $\mathbf{X}$. It can be clearly seen that the nodes in $\mathbf{Y}$ remain in yellow, which means that $P(\mathbf{y}|\mathbf{z}) = P(\mathbf{y}|\mathbf{z},\mathbf{x})$.

IV. EXPLANATION OF INFLUENCE DIAGRAMS IN ELVIRA

A. Explanation of the model

The explanation of IDs in Elvira is based, to a great extent, on the methods developed for explanation of BNs. One of the methods that have proven to be more useful is the automatic colorings of links. The definitions in Section III-A.2 for the sign of influence and magnitude of influence, inspired on [22], have been adapted to utility nodes as follows:

Definition 8: Let U be an ordinary utility node having $\alpha_U \neq \beta_U$ (see Equations 13 and 14) and $Pa(U) = \{A\} \cup \mathbf{B}$. The *magnitude of the influence* (MI) for the link $A \rightarrow U$ is

$$MI(A,U) = norm_U(\max_{a,\mathbf{b}} |\psi_U(a,\mathbf{b}) - \psi_U(a_0,\mathbf{b})|) \quad (29)$$

We say that A *positively influences* variable U iff $MI(A,U) \neq 0$ and

$$\forall a, \forall a', \forall \mathbf{b}, a > a' \Longrightarrow \psi_U(a,\mathbf{b}) \geq \psi_U(a',\mathbf{b}) \quad (30)$$

We also say that the link is *positive*.

The definitions of *negative influence* and *negative link* are analogous. When $MI(A,U) = 0$ the influence of link $A \rightarrow U$ is said to be *null*; in that case, link $A \rightarrow U$ should be removed. When the influence is neither positive nor negative nor null, then it is said to be *undefined*.

For instance, in Fig. 1 the link $X \rightarrow Y$ is colored in red because it represents a positive influence: the presence of the disease increases the probability of a positive result of the test. The link $X \rightarrow U_1$ is colored in blue because it represents a negative influence: the disease decreases the expected quality of life. The link $D \rightarrow U_1$ is colored in purple because its influence is undefined: the treatment is beneficial for patients suffering from X but detrimental for healthy patients.

As in the case of BNs, the coloring of links in Elvira has been very useful for debugging IDs, by detecting probability and utility tables whose numerical values do not agree with the qualitative influences assessed by the expert.

B. Displaying the results of inference

In Section II-B.3 we have seen that, given a strategy Δ, an ID can be converted into a Cooper policy network (CPN), which is a true Bayesian network. Consequently, all the explanation capabilities for BNs are also available for IDs by exploiting such transformation.

The information displayed for nodes depends on the kind of node—see Fig. 6. Chance and decision nodes display bars and numbers corresponding to the probabilities of their states, $P_\Delta(v)$, a marginal probability of $P_\Delta(\mathbf{v}_C, \mathbf{v}_D)$, defined by Equation 5. $P_\Delta(v)$ is the probability that a chance variable V takes a certain value v, or the probability that the decision maker chooses option v for decision V [32]. $P_\Delta(v)$ can be computed on the Cooper policy network (CPN) by means of Equation 16. Each utility node U displays the expected utility $EU_U(\Delta)$, defined by Equation 8, which is computed by propagating on the CPN and transforming back with the use of Equation 18. The guide bar (black line) indicates the range of the utilities.

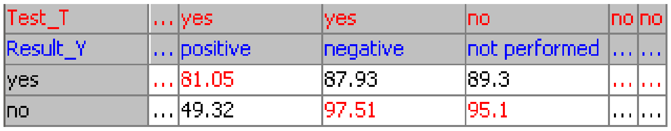
Links pointing into a decision node D are drawn with the color and thickness indicated in Section III-A.2, by examining the policy P_D (returned by the evaluation of the ID) as if it were the conditional probability table of a chance node. Non-forgetting links added during the evaluation of the diagram [33], [34], such as link $T \rightarrow D$ in Fig. 6, are drawn as discontinuous arrows.

Elvira, as most software tools for IDs, can show the utility table associated to each decision. For instance, in Table II each column corresponds to a configuration (t, y)

of the informational predecessors of decision D and each cell contains the expected utility of option d given t and y provided that every future decision will be made optimally $EU(d|\textit{iPred}(d)) = EU(d|t,y)$. In that table the order of the variables in *IPred(D)* is chosen to make it compatible with the partial order induced by the ID, i.e., the order in which the observations and decisions are known by the decision maker during the decision process.

TABLE II

EXPECTED UTILITIES FOR DECISION D IN THE ID IN FIGURE 1.



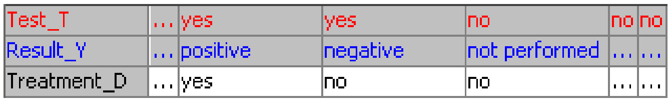
Test_T	...	yes	yes	no	no	no
Result_Y	...	positive	negative	not performed	...	...
yes	...	81.05	87.93	89.3	...	...
no	...	49.32	97.51	95.1	...	...

The highest utility in each column is highlighted in red. We have contracted the columns that represent impossible scenarios, i.e., configurations such that $P(\textit{iPred}(D))$=0.

This table is used by the evaluation algorithm to compute the optimal policy; in this example, $d_{opt} = \arg\max_d \ EU(d|t,y)$, as shown in Table III. A toggle allows the user to view either the expected utilities for a decision (Table II) or the optimal policy (Table III).

TABLE III

OPTIMAL POLICY FOR DECISION D IN THE ID IN FIGURE 1.



Test_T	...	yes	yes	no	no	no
Result_Y	...	positive	negative	not performed	...	...
Treatment_D	...	yes	no	no	...	...

C. Explanation of reasoning: decision trees

Initially, IDs were proposed as an alternative representation for decision trees (DTs) [2]. Not surprisingly, the first algorithm for evaluating IDs was to expand the equivalent DTs. Nowadays we have much more efficient algorithms for IDs, such as arc reversal [17], [33], [34] and variable elimination [3], [35], but the operations that they perform are only understood by experts in probabilistic graphical models. On the contrary DTs are easily understood by many users, because human beings tend to analyze decision problems by figuring out the possible scenarios, and each branch of a DT just represents a possible scenario, having a certain probability and a certain utility. An additional reason for using DTs when building medical decision-support systems is that most of the physicians learned about them as pregraduate students, and even many of them have done decision analysis with some software packages for DTs.

For this reason, even though Elvira uses the most efficient algorithms for evaluating IDs (otherwise it would be impossible to solve large models), it also offers the possibility of converting an ID into an equivalent DT and expanding and contracting its branches to the desired level of detail. Clearly, in the case of models containing dozens of nodes only a fraction of the branches can be expanded.

This idea, even though not original, has proven to be very useful in many situations. For instance, given the ID in Fig. 1, if the user wonders how Elvira obtained the utilities and the policy for D, it is possible to expand the DT shown in Fig. 7. In particular, the value $EU(D{=}\textsf{yes}|T{=}\textsf{yes}, Y{=}\textsf{positive}) = 81.05$ in Table II, which also appears in the branch $\{T{=}\textsf{yes}, Y{=}\textsf{positive}, D{=}\textsf{yes}\}$ in the DT, can be explained as the weighted average of the utility for the presence of the disease ($U = 78.00$, with probability 0.70) and the utility for the absence of the disease ($U = 88.00$, with probability 0.30). In turn, the utility for the disease, $U = 78.00$ can be explained as the utility associated to the quality of life, $U_1(x,d) = 80.00$ minus the cost of test, $U_2 = 2.00$. In the same way, the DT can explain the value $EU(D{=}\textsf{no}|T{=}\textsf{yes}, Y{=}\textsf{positive}) = 49.32$ in Table II.

The optimal decision for scenario $\{T{=}\textsf{yes}, Y{=}\textsf{positive}\}$ is $D{=}\textsf{yes}$, because $81.05 > 49.32$. For this reason, branch $\{T{=}\textsf{yes}, Y{=}\textsf{positive}, D{=}\textsf{yes}\}$ in the DT is highlighted with a red square, in accordance the highlighting of value 81.05 in Table II.

Therefore, the main difference of Elvira with respect to other software tools is that, in addition to showing the (global) expected utility of each branch, it can also show the individual utilities that compose it, i.e., the utilities associated to the utility nodes other than U_0.

D. Introduction of evidence

Elvira's ability to manage several evidence cases simultaneously in BNs is also available for IDs. The evidence is introduced in the ID by using its corresponding Cooper policy network. Given evidence $\mathbf{e}$, Elvira displays for each chance and decision node V the probability $P_\Delta(v|\mathbf{e})$ (cf. Eqs. 6 and 17), and for each utility node U the expected utility $EU_U(\Delta, \mathbf{e})$ (cf. Eqs. 7 and 19), as shown in Fig. 8.

1) Clarifying the concept of evidence in influence diagrams: In order to avoid confusions, we must mention that the meaning of evidence in Elvira is very different from its meaning in some methods oriented to the computation of the value of information in IDs, such as [36], [37], [38]. For those methods, the introduction of evidence $\mathbf{e}$ leads to a different decision problem in which the values of the variables in $\mathbf{E}$ would be known with certainty before making any decision. For instance, introducing evidence $\{+x\}$ in the ID in Fig. 1 would mean that X were known when making decisions T and D. Therefore, the expected utility of the new decision problem, which we call "Ezawa's scenario" [38], would be

$$\max_t \sum_y \max_d P(y|+x:t,d) \cdot \underbrace{(U_1(+x,d) + U_2(t))}_{U_0(+x,d,t)}$$

where $P(y|+x : t,d) = P(+x,y : t,d)/P(+x : t,d) = P(+x,y : t)/P(+x) = P(y|+x : t)$. In spite of the apparent similarity of this expression with Equation 12, the optimal strategy changes significantly from "test, and treat only if the result is positive" to "always treat, without testing", because if we knew with certainty that the disease X is present. the result of the test would be irrelevant. The *MEU* for the new decision problem would be $U_0(+x, +d, \neg t) = U_1(+x, +d)$.

In contrast, the introduction of evidence in Elvira does not lead to a new decision scenario nor to a different strategy, since the strategy is determined *before* introducing the "evidence". Put another way, when introducing evidence in Elvira we adopt the point of view of an external observer of a system including the decision maker as one of its components. The probabilities and expected utilities given by Equations 5 and 7 are those corresponding to the subpopulation indicated by $\mathbf{e}$ when the decision maker applies strategy Δ. For instance, given the evidence $\{+x\}$, the probability $P_\Delta(+t|+x)$ shown by Elvira is the probability that a patient suffering from X receives the test, which is 1 (it was 0 in Ezawa's scenario), and $P_\Delta(+d|+x)$ is the probability that he receives the treatment; contrary to Ezawa's scenario, this probability may differ from 1 because of false negatives. The expected utility for a patient suffering from X is

$$\begin{aligned} & EU(\Delta, \{+x\}) = \\ & = \sum_{t,y,d} P_\Delta(t, y, d|+x) \cdot (U_1(+x, d) + U_2(t)) \end{aligned}$$

where $P_\Delta(t, y, d|+x) = P_\Delta(t) \cdot P(y|t, +x) \cdot P_\Delta(d|t, y)$. For the optimal strategy,

$$\begin{aligned} EU(\Delta_{opt}, \{+x\}) =& [P(+y|+x) \cdot U_1(+x, +d) \\ & + P(\neg y|+x) \cdot U_1(+x, \neg d)] + U_2(+t) \end{aligned}$$

A second difference is that the evidence introduced in Elvira may include "findings" for decision variables. For instance, $\mathbf{e} = \{+d\}$ would represent the subpopulation of patients who have received therapy, and $P_\Delta(+x|+d)$ is the probability that a patient receiving therapy has disease X.

And the third difference is that Elvira admits the possibility of analyzing non-optimal strategies, as we will see below.

We must stress that the two approaches are not rivals. They correspond to different points of view when considering evidence in IDs and can complement each other in order to perform a better decision analysis and to explain the reasoning. We have implemented first the options that, in our opinion, can be more useful, but in the future we will implement as well Ezawa's method and the possibility of computing the expected value of perfect information (EVPI).

2) Example: Fig. 8 shows two evidence cases. In this example, Δ is the optimal strategy obtained when evaluating the ID, because no policy was imposed by the user. The first evidence case in Fig. 8 is the *prior case*, which was also displayed in Fig. 6. Its probabilities and expected utilities are those of the general population. The second evidence case is given by $\mathbf{e} = \{+y\}$; i.e., it displays the probabilities and utilities of the subpopulation of patients in which the test has given a positive result. Node Y is colored in gray to highlight the fact that there is evidence about it. The probability $P_\Delta(+x|+y)$, represented by a red bar, is 0.70; the green bar close to it represents the probability of $+x$ for the prior case, i.e., $P_\Delta(+x)$, which equals $P(+x)$ because the decision maker's actions do not affect X. The red bar is longer than the green one because $P_\Delta(+x|+y) > P_\Delta(+x)$, as it was expected from the fact that link $X \rightarrow Y$ is positive. The global utility for the second evidence case, $EU(\Delta, \{+y\})$, represented by a red bar in node U_0, is smaller than $EU(\Delta, \varnothing)$, the expected utility for the general population, represented by a green bar, because the presence of the symptom worsens the prognosis. The red bar for Treatment=yes, which represents $P_\Delta(+d|+y)$, is 1.00 because the optimal strategy determines that all symptomatic patients must be treated. Similarly, $P_\Delta(+t|+y) = 1.00$ because a positive result of the test implies that the test has been done.

3) Debugging influence diagrams by introducing evidence: The possibility of introducing evidence in Elvira has been useful for building IDs in medicine [10]: before having this explanation facility, when we were interested in computing the posterior probability of a certain diagnosis given a set of findings, we needed to manually convert the ID into a BN by removing decision and utility nodes. Each time the ID was modified, even slightly, we had to repeat this conversion, which was tedious and time consuming. (When building medical expert systems, the time of interaction with the experts is a precious resource that must not be waisted.) This was the reason for implementing a facility that allowed us to compute the probabilities directly on the ID, which is much more convenient.

E. What-if reasoning: analysis of non-optimal strategies

In Elvira it is possible to have a strategy in which some of the policies are imposed by the user and the others are computed by maximization. The way of imposing a policy consists in setting a probability distribution P_D for the corresponding decision D by means of Elvira's GUI; the process is identical to editing the conditional probability table of a chance node. In fact, such a decision will be treated by the inference algorithms as if it were a chance node, and the maximization will be performed only for the rest of the decisions.

This way, in addition to computing the optimal strategy (when the user has imposed no policy), as any other software tool for IDs, Elvira also permits to analyze how the expected utilities and the rest of the policies would vary if the decision maker chose a non-optimal policy for some of the decisions (what-if reasoning).

The reason for implementing this explanation facility is that when we were building a certain medical influence diagram [10] our expert wondered why the model recommended not to perform a certain test. We wished to compute the a posteriori probability of the disease given a positive result in the test, $P_\Delta(+x|+y)$, but we could not introduce this "evidence", because it was incompatible with the optimal policy (not to test): $P_\Delta(+y) = 0$. After we implemented the possibility of imposing non-optimal policies (in this case, performing the test) we could see that the posterior probability of the disease remained below the treatment threshold even after a positive result in the test, and given that the result of the test would be irrelevant, it was not worthy to do it.

F. Sensitivity analysis and decision thresholds

Recently Elvira has been endowed with some well-known sensitivity analysis tools, such as one-way sensitivity analysis, tornado diagrams, and spider diagrams [39], which can be combined with the above-mentioned methods for the explanation of reasoning. One-way sensitivity analysis can be used for finding treatment thresholds in different scenarios [40] and, in consequence, to explain the optimal policies. For instance, in Fig. 9, which shows the results of one-way sensitivity analysis on the prevalence of X for the ID given in Fig. 1. This graph is obtained by evaluating several instances of the ID, each having a different value of $P(+x)$. We can see that the treatment threshold is approximately 0.17, i.e., when $P(+x) < 0.17$ the best option is not to treat the patient, and when $P(+x) > 0.17$ it is better to treat.

By introducing evidence about Y in the ID we can see that $P(+x|+y) = 0.83$; this means that the prevalence of X in the subpopulation $\{+y\}$ is 0.83, which is above the 0.17 threshold. In contrast, $P(+x|\neg y) = 0.015 < 0.17$. This explains why the optimal policy for D is to treat only after a positive result of the test. In the construction of more complex IDs this kind of analysis has been useful for understanding why some tests are necessary or not, and why sometimes the result of a test is irrelevant, as discussed in the previous section.

V. RELATED WORK AND FUTURE RESEARCH

In spite of the importance of explanation in artificial intelligence, for the reasons mentioned in Section I, most software tools for building expert systems —either probabilistic or heuristic— offer no facilities for this task. There are several prototype systems developed to demonstrate new explanation options, but most of them never became publicly available—see [4], [7] for a review. Among the few available environments for probabilistic graphical models endowed with some explanation facilities, we can mention the following:

- BayesiaLab[7], Netica[8], and especially SamIam[9], are able to perform an *analysis of sensitivity to the parameters* of the model in BNs (see Footnote 6). TreeAge[10] can convert an ID into a decision tree (in fact, the main representation framework in TreeAge are decision trees, not IDs) and to perform several types of analysis of sensitivity to the parameters.
- GeNIE[11], Hugin[12], and MSBNx[13] permit to compute the *value of information*.
- Recent versions of BayesiaLab are able to simultaneously display probability bars for several evidence cases simultaneously, and to represent the sign and magnitude of influences by the color and thickness of links, as in Elvira.
- The latest version of Hugin can perform an *analysis of sensitivity to evidence* in BNs and IDs.

Clearly, Elvira offers more explanation facilities than any of these tools, but there are still many options that can be added. In particular, we will further explore the generation of verbal explanations and the possibility of adapting them to different user needs. However, the goal of offering a natural language dialog between human users and Elvira is still impossible given the current state of the art.

We are currently in the process of implementing probabilistic analysis of sensitivity to the parameters in Elvira. Other kinds of sensitivity analysis for both BNs and IDs will be added in the future. It would also be possible to integrate sensitivity analysis in IDs with the expansion of decision trees.

As mentioned above, we also intend to implement new facilities for introducing evidence in Ezawa's style [38] and for computing the value of information—see Sec. IV-D.1. On the other hand, we are studying how to obtain a set of rules that summarize a strategy; for instance, "if the patient presents with symptom S and the blood test is positive, then apply treatment T; otherwise, do not treat"; see the work by Fernández del Pozo et al. [41] on this subject. The work on explanation in factored Markov decision processes by Elizalde et al. [42], which focuses on the variable that receives the higher impact from an action performed by the decision maker, may be applied to ordinary IDs as well.

Other techniques indirectly related with explanation are the facility of dealing with submodels, the possibility of simplifying the model [5] in order to improve both the efficiency of the algorithms and the generation of explanations, the application of different techniques for the visualization of Bayesian networks [43], and the development of a graphical user interface that facilitates the construction of BNs and IDs by non-expert users.

In any case, the new explanation options for Elvira will be selected as a response to the most relevant needs detected in our research on building medical application and in our task of teaching probabilistic graphical models to computer science students and to health professionals.

Finally, we intend to use Elvira for developing models in other domains, such as financial risk analysis, cost-effectiveness studies, and collaborative e-learning, which will certainly pose new challenges for the explanation of the models and the reasoning.

VI. CONCLUSION

In this paper we have described the main explanation facilities for Bayesian networks and influence diagrams implemented in Elvira, a public software package, and how they have helped us in building medical applications [23] and when teaching probabilistic graphical models to pre- and postgraduate students [11]. In our experience, the most useful explanation options for BNs among those offered by Elvira are, in this order, the simultaneous display of several evidence cases, the possibility of coding the sign and magnitude of influences by the color and thickness of links, and the explanation of evidence cases by highlighting the active chains

[7] http://www.bayesia.com.

[8] http://www.norsys.com/netica.html.

[9] http://reasoning.cs.ucla.edu/samiam/. Its acronym stands for "Sensitivity Analysis, Modeling, Inference and More".

[10] http://www.treeage.com/.

[11] http://genie.sis.pitt.edu/.

[12] http://www.hugin.com.

[13] http://research.microsoft.com/adapt/MSBNx/.

of reasoning, which includes the coloring of nodes in those chains. With respect to IDs, the most useful options are the possibility of introducing evidence, the conversion of IDs into decision trees, the possibility of analyzing non-optimal policies imposed by the user, and the analysis of sensitivity to the parameters. Further research is still necessary to make probabilistic reasoning more understandable to human users.

Acknowledgments

The conversion of IDs into decision trees and sensitivity analysis in IDs have been implemented in Elvira by Jorge Fernández and Jacinto Ruiz, respectively, two students of the Computer Science School of the UNED, under the supervision of the authors of this paper.

This work has been partially financed by the Spanish CICYT under projects TIC-97-1135-C04 and TIC-2001-2973-C05. The second author has benefited from a grant from the Department of Education of the Comunidad de Madrid, partially covered by the European Social Fund (ESF).

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[2] R. A. Howard and J. E. Matheson, "Influence diagrams," in *Readings on the Principles and Applications of Decision Analysis*, R. A. Howard and J. E. Matheson, Eds. Menlo Park, CA: Strategic Decisions Group, 1984, pp. 719–762.

[3] F. V. Jensen, *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2001.

[4] C. Lacave and F. J. Díez, "A review of explanation methods for Bayesian networks," *Knowledge Engineering Review*, vol. 17, pp. 107–127, 2002.

[5] C. Lacave, "Explanation in causal Bayesian networks. Medical applications," Ph.D. dissertation, Dept. Inteligencia Artificial. UNED, Madrid, Spain, 2003, in Spanish.

[6] J. W. Wallis and E. H. Shortliffe, "Customized explanations using causal knowledge," in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B. G. Buchanan and E. H. Shortliffe, Eds. Reading, MA: Addison-Wesley, 1984, ch. 20, pp. 371–388.

[7] C. Lacave and F. J. Díez, "A review of explanation methods for heuristic expert systems," *Knowledge Engineering Review*, vol. 19, pp. 133–146, 2004.

[8] F. J. Díez, J. Mira, E. Iturralde, and S. Zubillaga, "DIAVAL, a Bayesian expert system for echocardiography," *Artificial Intelligence in Medicine*, vol. 10, pp. 59–73, 1997.

[9] C. Lacave and F. J. Díez, "Knowledge acquisition in Prostanet, a Bayesian network for diagnosing prostate cancer," in *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES'2003)*, ser. Lecture Notes in Computer Science, vol. 2774. Oxford, UK: Springer, Berlin, Germany, 2003, pp. 1345–1350.

[10] M. Luque, F. J. Díez, and C. Disdier, "Influence diagrams for medical decision problems: Some limitations and proposed solutions," in *Proceedings of the Intelligent Data Analysis in Medicine and Pharmacology*, J. H. Holmes and N. Peek, Eds., 2005, pp. 85–86.

[11] F. J. Díez, "Teaching probabilistic medical reasoning with the Elvira software," in *IMIA Yearbook of Medical Informatics*, R. Haux and C. Kulikowski, Eds. Stuttgart: Schattauer, 2004, pp. 175–180.

[12] M. Henrion and M. Druzdzel, "Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning," in *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence (UAI'90)*, Cambridge, MA, 1990, pp. 17–32.

[13] P. Sember and I. Zukerman, "Strategies for generating micro explanations for Bayesian belief networks," in *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, Windsor, Ontario, 1989, pp. 295–302.

[14] T. Elvira Consortium, "Elvira: An environment for creating and using probabilistic graphical models," in *Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM'02)*, Cuenca, Spain, 2002, pp. 1–11.

[15] C. Lacave, R. Atienza, and F. J. Díez, "Graphical explanation in Bayesian networks," in *Proceedings of the International Symposium on Medical Data Analysis (ISMDA-2000).* Frankfurt, Germany: Springer-Verlag, Heidelberg, 2000, pp. 122–129.

[16] M. J. Druzdzel and R. Flynn, "Decision support systems," in *Encyclopedia of Library and Information Science*, A. Kent, Ed. New York: M. Dekker, Inc., 2000, vol. 67, pp. 120–133.

[17] J. A. Tatman and R. D. Shachter, "Dynamic programming and influence diagrams," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, pp. 365–379, 1990.

[18] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag, 1999.

[19] G. F. Cooper, "A method for using belief networks as influence diagrams," in *Proceedings of the 4th Workshop on Uncertainty in AI*, University of Minnesota, Minneapolis, MN, 1988, pp. 55–63.

[20] V. L. Yu, L. M. Fagan, S. M. Wraith, W. J. Clancey, A. C. Scott, J. F. Hannigan, R. L. Blum, B. G. Buchanan, and S. N. Cohen, "An evaluation of MYCIN's advice," in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B. G. Buchanan and E. H. Shortliffe, Eds. Reading, MA: Addison-Wesley, 1984, ch. 31, pp. 589–596.

[21] A. Oniśko, M. J. Druzdzel, and H. Wasyluk, "Extension of the Hepar II model to multiple-disorder diagnosis," in *Intelligent Information Systems*, M. Kłopotek, M. Michalewicz, and S. Wierzchoń, Eds. Springer-Verlag, Heidelberg, 2000, pp. 303–313.

[22] M. P. Wellman, "Fundamental concepts of qualitative probabilistic networks," *Artificial Intelligence*, vol. 44, pp. 257–303, 1990.

[23] C. Lacave, A. Oniśko, and F. J. Díez, "Use of Elvira's explanation facilities for debugging probabilistic expert systems," *Knowledge-Based Systems*, pp. 730–738, 2006.

[24] H. J. Suermondt, "Explanation in Bayesian belief networks," Ph.D. dissertation, Dept. Computer Science, Stanford University, STAN–CS–92–1417, 1992.

[25] F. V. Jensen, B. Chamberlain, T. Nordahl, and F. Jensen, "Analysis in HUGIN of data conflict," in *Uncertainty in Artificial Intelligence 6*, P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds. Amsterdam, The Netherlands: Elsevier Science Publishers, 1991, pp. 519–528.

[26] F. V. Jensen, S. H. Aldenryd, and K. B. Jensen, "Sensitivity analysis in bayesian networks," *Lecture Notes in Artificial Intelligence*, vol. 946, pp. 243–250, 1995.

[27] E. Castillo, J. M. Gutiérrez, and A. S. Hadi, "Sensitivity analysis in discrete Bayesian networks," *IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans*, vol. 27, pp. 412–423, 1997.

[28] H. Chan and A. Darwiche, "When do numbers really matter?" *Journal of Artificial Intelligence Research*, vol. 17, pp. 265– 287, 2002.

[29] V. M. H. Coupé and L. C. van der Gaag, "Properties of sensitivity analysis of Bayesian belief networks," *Annals of Mathematics and Artificial Intelligence*, vol. 36, pp. 323–356, 2002.

[30] K. B. Laskey, "Sensitivity analysis for probability assessments in Bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 901–909, 1995.

[31] B. G. Buchanan and E. H. Shortliffe, Eds., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley, 1984.

[32] D. Nilsson and F. V. Jensen, "Probabilities of future decisions," in *Proceedings from the International Conference on Informational Processing and Management of Uncertainty in knowledge-based Systems*, 1998, pp. 1454–1461.

[33] S. M. Olmsted, "On representing and solving decision problems," Ph.D. dissertation, Dept. Engineering-Economic Systems, Stanford University, CA, 1983.

[34] R. D. Shachter, "Evaluating influence diagrams," *Operations Research*, vol. 34, pp. 871–882, 1986.

[35] M. Luque and F. J. Díez, "Variable elimination for influence diagrams with super-value nodes," in *Proceedings of the Second European Workshop on Probabilistic Graphical Models*, P. Lucas, Ed., 2004, pp. 145–152.

[36] R. Shachter and M. Peot, "Decision making using probabilistic inference methods," in *Proceedings of the 8th Annual Conference on Uncertainty in Artificial Intelligence (UAI-92)*. San Mateo, CA: Morgan Kaufmann, 1992, pp. 276–28.

[37] S. L. Dittmer and F. V. Jensen, "Myopic value of information in influence diagrams," in *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI'97)*. Providence, RI: Morgan Kaufmann, San Francisco, CA, 1997, pp. 142–149.

[38] K. Ezawa, "Evidence propagation and value of evidence on influence diagrams," *Operations Research*, vol. 46, pp. 73–83, 1998.

[39] R. T. Clemen and T. A. Reilly, *Making Hard Decisions*. Pacific Grove, CA: Duxbury, 2001.

[40] L. C. van der Gaag and V. M. Coupe, "Sensitivity analysis for threshold decision making with Bayesian belief networks," *Lecture Notes in Artificial Intelligence*, vol. 1792, pp. 37–48, 2000.

[41] J. A. Fernández del Pozo, C. Bielza, and M. Gómez, "A list-based compact representation for large decision tables management," *European Journal of Operational Research*, vol. 160, pp. 638–662, 2005.

[42] F. Elizalde, E. Sucar, and P. de Buen, "Explanation generation through probabilistic models for an intelligent assistant," Submitted to *IBERAMIA'06*, 2006.

[43] J. D. Zapata-Rivera, E. Neufeld, and J. Greer, "Visualization of Bayesian belief networks," in *IEEE Visualization 1999. Late Breaking Hot Topics Proceedings*, San Francisco, CA, 1999, pp. 85–88.

Carmen Lacave was born in Valdepeas (Ciudad Real), Spain, in 1967. She received a M.S. in Mathematics from the Universidad Complutense de Madrid in 1990 and a Ph.D. from the Universidad Nacional de Educacin a Distancia (UNED) in 2003. She is currently professor at the Department of Technologies and Systems of Information at the Universidad de Castilla-La Mancha, in Ciudad Real, and member of the Research Center on Intelligent Decision-Support Systems (CISIAD), at the UNED.

Manuel Luque was born in Madrid, Spain, in 1977. He received a Bs.D. in Computer Science in 2003, at the University of Mlaga, Spain. He is a Ph.D. candidate in the decision-support systems unit of the Department of Artificial Intelligence, at the Universidad Nacional de Educacin a Distancia (UNED), Madrid, since 2003. In the autumns of 2005 and 2006 he was a visiting scholar at the Machine Intelligence Group, which is part of the Department of Computer Science of Aalborg University, in Denmark.

Francisco Javier Dez was born in Burgos, Spain, in 1965. He received a M.S. in Theoretical Physics from the Universidad Autnoma de Madrid and a Ph.D. from the Universidad Nacional de Eduacin a Distancia (UNED) in 1994. He is currently associate professor at the Department of Artificial Intelligence at the UNED, in Madrid, and Director of the Research Center on Intelligent Decision-Support Systems (CISIAD), at the same university.

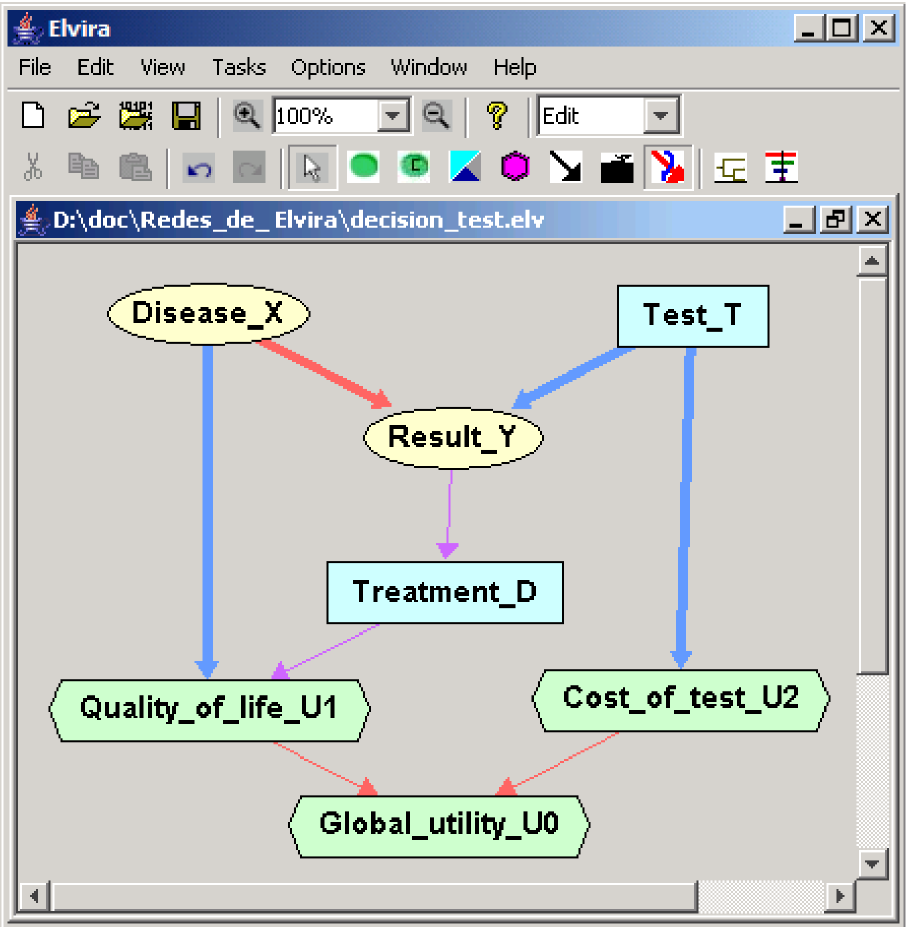


Fig. 1. ID with two decisions (rectangles), two chance nodes (ovals) and three utility nodes (hexagons). Please note that there is a directed path T–Y–D–U_1–U_0 including all the decisions and the global utility node U_0.

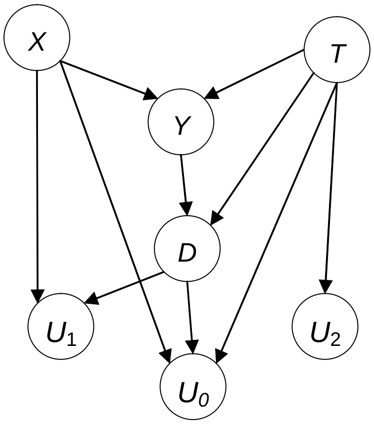


Fig. 2. Cooper policy network (CPN) for the ID in Figure 1. Please note the addition of the non-forgetting link $T \rightarrow D$ and that the parents of node U_0 are no longer U_1 and U_2 but $FPred(U_0) = \{X, D, T\}$, which were chance or decision nodes in the ID.

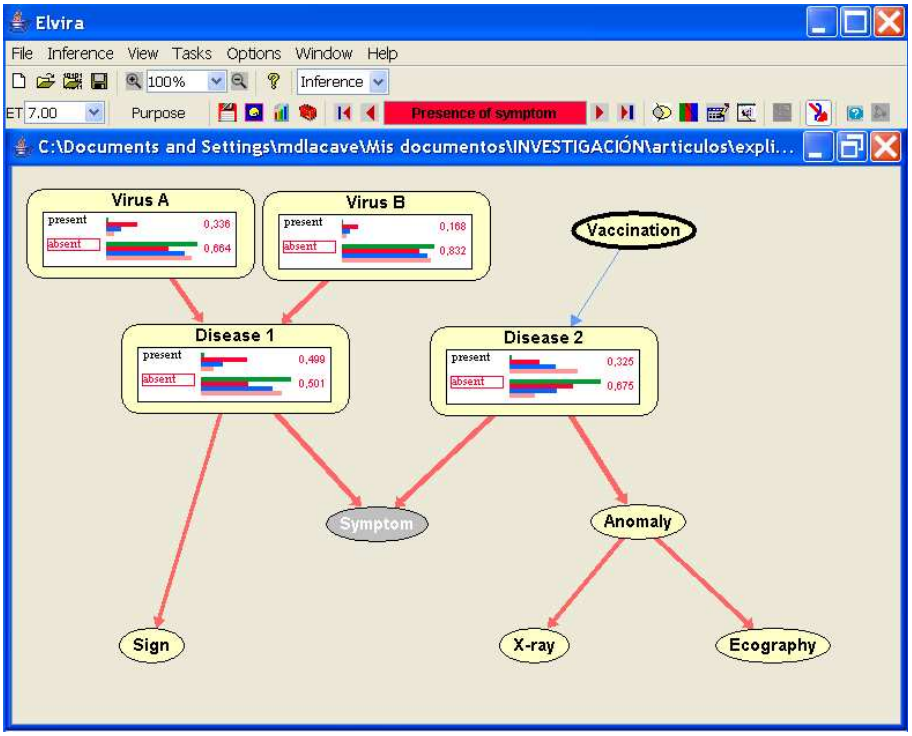


Fig. 3. Elvira main window in inference mode.

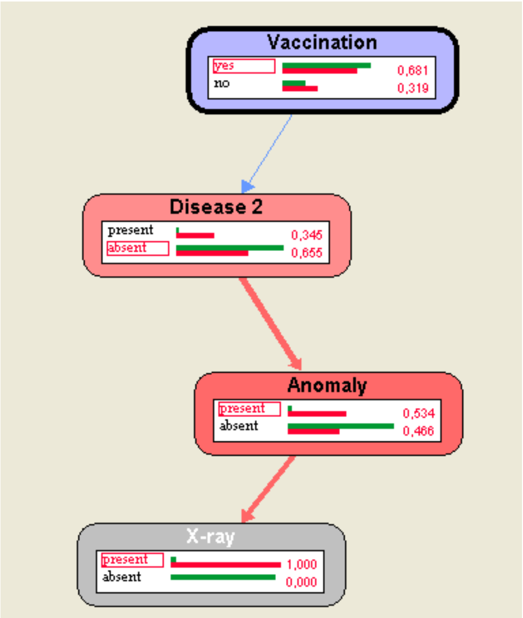
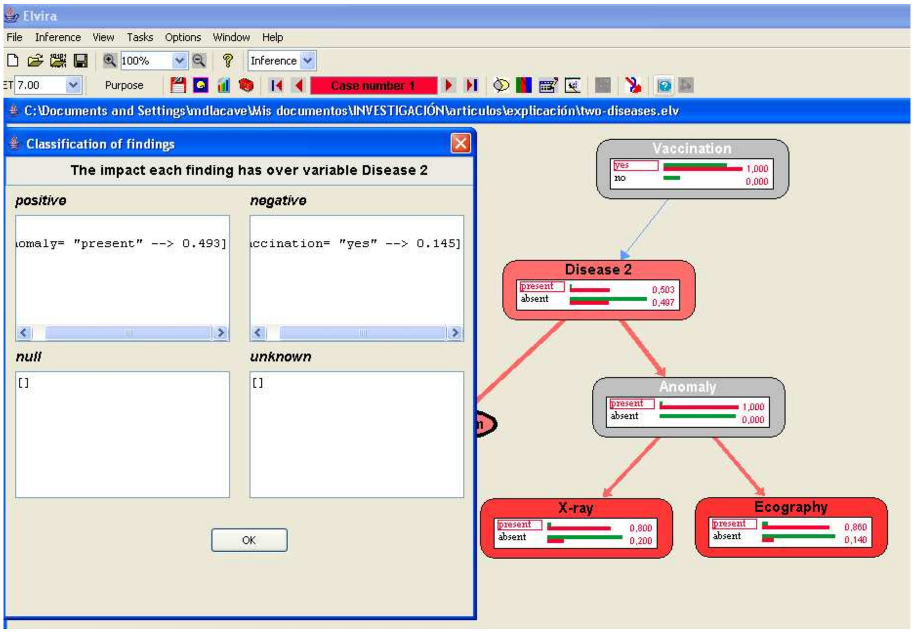


Fig. 4. Chains of reasoning for the graphical explanation of a case whose evidence is defined by the finding X-ray=positive. The selection of a variable of interest (Vaccination in this example) makes Elvira hide the nodes and links that do not make part of any active path from the evidence to the variable of interest.



Classification of findings

The impact each finding has over variable Disease 2

positive

```
omaly= "present" --> 0.493]
```

negative

```
ccination= "yes" --> 0.145]
```

null

```
[]
```

unknown

```
[]
```

OK

Fig. 5. This figure illustrates the impact that each finding, Vaccination=yes and Anomaly=present, has separately over the selected node Disease 2 in the network of Fig. 3.

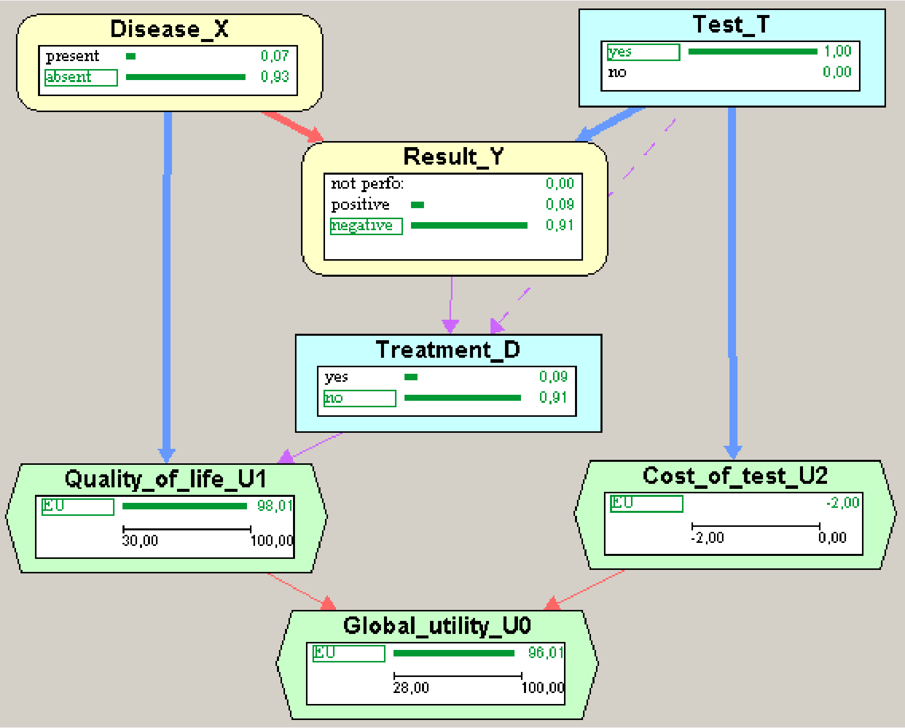
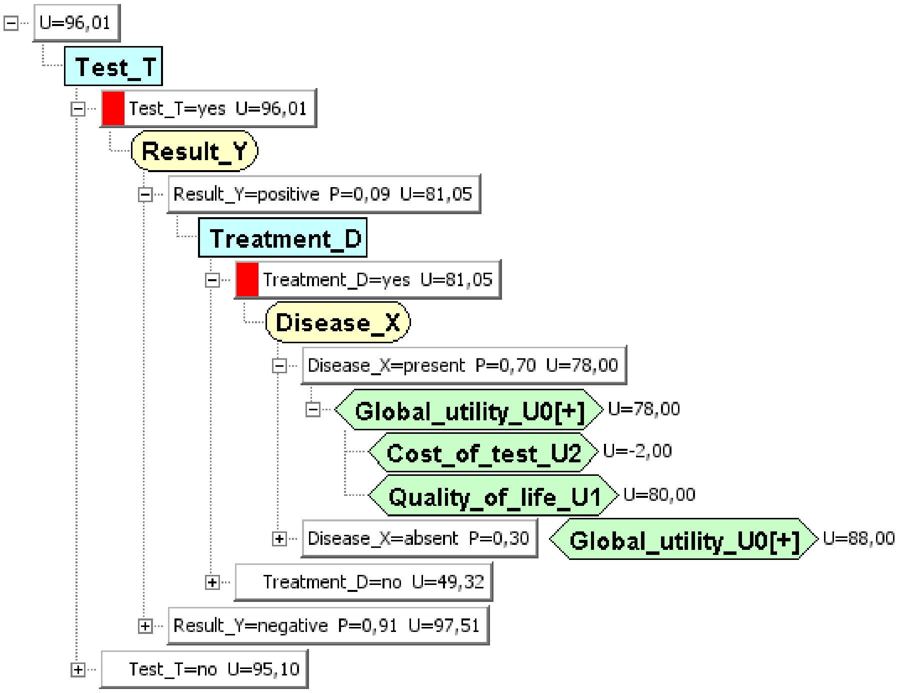


Fig. 6. ID resulting from the evaluation of the ID in Figure 1. It shows the probability $P_{\triangle}(v)$ of each chance and decision node and the expected utilities.



```
⊟ U=96,01
  Test_T
    ⊟ Test_T=yes  U=96,01
        Result_Y
          ⊟ Result_Y=positive  P=0,09  U=81,05
              Treatment_D
                ⊟ Treatment_D=yes  U=81,05
                    Disease_X
                      ⊟ Disease_X=present  P=0,70  U=78,00
                          ⊟ Global_utility_U0[+]  U=78,00
                              Cost_of_test_U2  U=-2,00
                              Quality_of_life_U1  U=80,00
                      ⊞ Disease_X=absent  P=0,30   Global_utility_U0[+]  U=88,00
                ⊞ Treatment_D=no  U=49,32
          ⊞ Result_Y=negative  P=0,91  U=97,51
    ⊞ Test_T=no  U=95,10
```

Fig. 7. Decision tree for the ID in Figure 1, where some branches have been expanded to obtain more level of the detail.

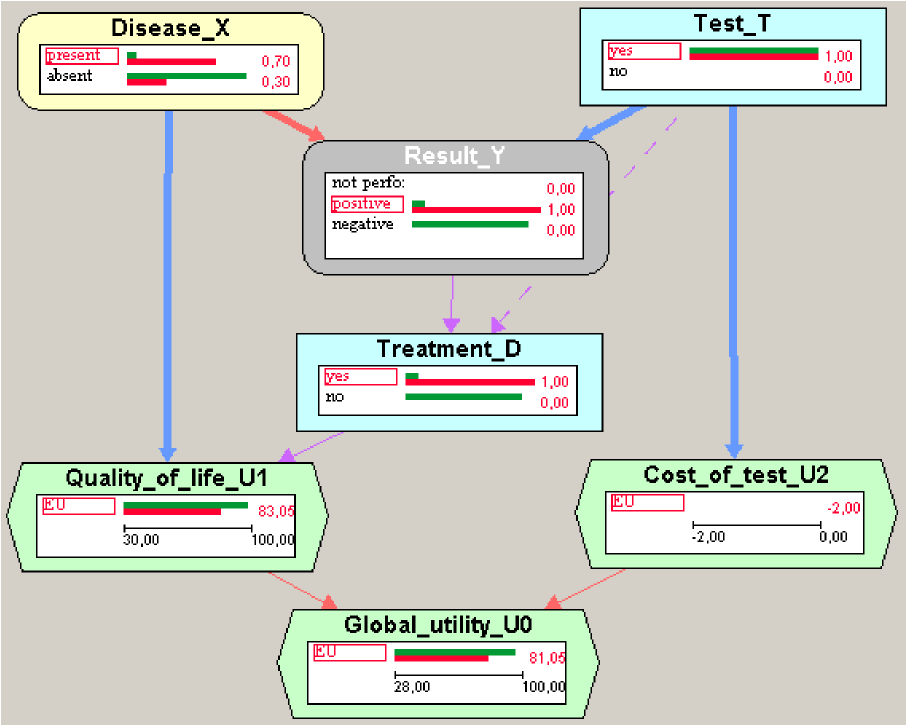


Fig. 8. ID resulting from the evaluation of the ID in Figure 1. It shows two evidence cases: the prior case (no evidence) and the case in which $\mathbf{e} = \{+y\}$.

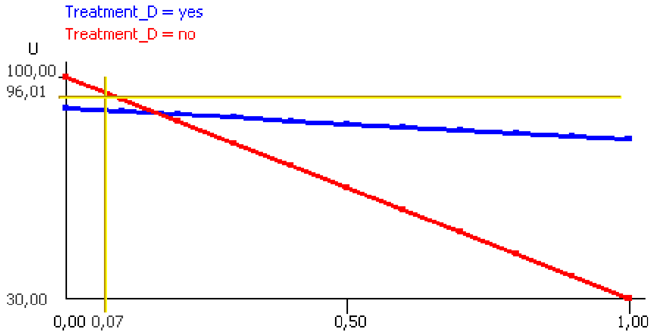


Fig. 9. Elvira's one-way sensitivity analysis on the prevalence of the disease, which is represented in the the x-axis. The y-axis represents the expected utility. The treatment threshold is 0.17.