

**SCHOOL OF BUSINESS WORKING PAPER NO. 319**

## Modeling Challenges with Influence Diagrams: Representation Issues

**Concha Bielza**

Dept. de Inteligencia Artificial, Universidad Politécnica de Madrid, Boadilla del Monte, Madrid 28660, SPAIN

[mbielza@fi.upm.es](mailto:mbielza@fi.upm.es)

**Manuel Gómez**

Dept. of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, SPAIN

[mgomez@decsai.ugr.es](mailto:mgomez@decsai.ugr.es)

**Prakash P. Shenoy**

School of Business, University of Kansas, Lawrence, KS 66045, USA

[pshenoy@ku.edu](mailto:pshenoy@ku.edu)

May 2009<sup>†</sup>

---

<sup>†</sup> Comments and suggestions for improvement are welcome and will be gratefully appreciated.

# Modeling Challenges with Influence Diagrams: Representation Issues

Concha Bielza

Dept. de Inteligencia Artificial, Universidad Politécnica de Madrid, Boadilla del Monte, Madrid 28660, SPAIN,  
mcbielza@fi.upm.es

Manuel Gómez

Dept. of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, SPAIN,  
mgomez@decsai.ugr.es

Prakash P. Shenoy

School of Business, University of Kansas, Lawrence, KS 66045, USA, pshenoy@ku.edu

Since their introduction in the mid 70s, influence diagrams have become a de facto standard for representation of Bayesian decision problems. The need to represent complex problems has led to extensions of the influence diagram methodology to increase their representation power. In this paper, we review the representation issues associated with influence diagrams. In particular, we look at the representation of asymmetric decision problems including conditional distribution trees, sequential decision diagrams, and sequential valuation networks. We also examine the issue of representing the sequence of decisions and uncertainties, and how it is done in unconstrained influence diagrams, sequential valuation networks, and sequential influence diagrams. We also analyze the use of continuous chance and decision variables, including continuous conditionally deterministic variables. Finally, we discuss some of the modeling challenges faced in representing decision problems in practice and relevant software.

*Key words:* Decision-making under uncertainty, influence diagrams, modeling, probabilistic graphical models, sequential decision diagrams, unconstrained influence diagrams, sequential valuation networks, sequential influence diagrams

---

## 1 Introduction

Influence diagrams (IDs) are graphical models for representing and solving complex decision-making problems based on uncertain information. Nowadays they have become a popular and standard modeling tool. As pointed out in a recent special issue of the *Decision Analysis* journal devoted to IDs, these models “command a unique position in the history of graphical models”, Pearl (2005). They were first used in 1973 by the Decision Analysis Group at Stanford Research Institute for a project for the Defense Intelligence Agency. IDs tried to model political conflicts in the Persian Gulf to see whether more intelligence resources should be allocated and they also tried to measure the value of information produced by the Agency. As chronicled by themselves in Howard et al. (2006), Miller invented a new

graphical convention that we now call ID, Howard coined the term “influence diagram” and the first resulting (classified) paper on IDs was North et al. (1974). But it is Howard and Matheson’s paper, Howard and Matheson (1984), reprinted in Howard and Matheson (2005b), what is considered to contain the invention of IDs and to be the pioneering paper. Other interesting details of the evolution and application of IDs since their inception may be found in a retrospective, Howard and Matheson (2005a).

IDs are directed acyclic graphs that may be seen as Bayesian networks augmented with decision and value nodes. Therefore, there are three types of nodes: (1) decision nodes (rectangular) representing decisions to be made; (2) chance nodes (oval or elliptical) representing uncertainties modelled by probability distributions; and (3) value nodes (diamond-shaped) without children (direct successors), representing the (expected) utilities that model decision-maker’s preferences. The arcs have different meanings depending upon which node they are directed to: the arcs to chance nodes or the value nodes indicate probabilistic dependence and functional dependence, respectively, while the arcs pointing at a decision node indicate the information known at the time of making that decision. The former are called *conditional* arcs while the latter are called *informational*.

At the qualitative (or graphical) level, the ID has a requirement that there must be a directed path comprising all decision nodes. It ensures that the ID defines a temporal sequence (total order) of decisions and it is called the *sequencing constraint*. As a consequence, IDs have the “*no-forgetting*” property: the decision maker remembers the past observations and decisions. In contrast, the transitive closure of the precedence binary relation induced by the informational arcs is a partial order on the set of all decision and chance nodes. Therefore, the information constraints represented by these arcs may not specify a complete order. From a semantic viewpoint, it means that the decision-maker may not impose constraints regarding which of two or more chance nodes must precede the others in the decision-making process. For example, after deciding the admission of a patient into a hospital, there will be no constraints concerning which of the uncertainties cost of stay and risks of being admitted (infections, contagions,...) is the first to occur.

At the quantitative level, an ID specifies the domains of all decision and chance nodes: the list of alternatives for each decision variable and the state space  $\Omega_X$  of each chance variable  $X$ . Also, a conditional probability table is attached to each chance node consisting of conditional probability distributions, one for each state of its parents (direct predecessors), and a real-valued function, the utility function, is defined over the states of a value node’s parents. Probabilities and utilities represent beliefs and preferences, respectively, of the decision maker. When a problem has a utility function that factorizes additively, each of the additive factors is represented as a value node.

Figure 1 shows an example of the graphical part of an ID.  $D_1, D_2$  are decisions,  $A, C, R$  are random

variables and  $v_1$ ,  $v_2$ , and  $v_3$  are value nodes.  $v_1$  is a function of the states of  $D_1$ ,  $v_2$  is a function of the states of  $D_2$  and  $A$ , and  $v_3$  is a function of the states of  $D_2$  and  $C$ . The joint utility function is the pointwise sum of  $v_1$ ,  $v_2$  and  $v_3$ . As in a Bayesian network, the arcs directed to chance nodes like  $R$  mean that the conditional probability attached to  $R$  is given by  $P(R|D_1, A)$ . In fact, the chain rule for this ID is:  $P(R, A, C|D_1) = P(R|D_1, A)P(A)P(C)$ . Finally, the informational arcs directed to  $D_2$  say that at the time  $D_2$  is made, we know the outcome of  $R$  and the decision made at  $D_1$ . Since we have a directed path  $D_1, R, D_2$ , the arc  $(D_1, D_2)$  is implied by the no-forgetting condition and could be omitted. These arcs imply that the information constraints specify the partial order given by (1)  $D_1$ , (2)  $R$ , (3)  $D_2$  and finally, (4)  $A$  or  $C$  in any order.

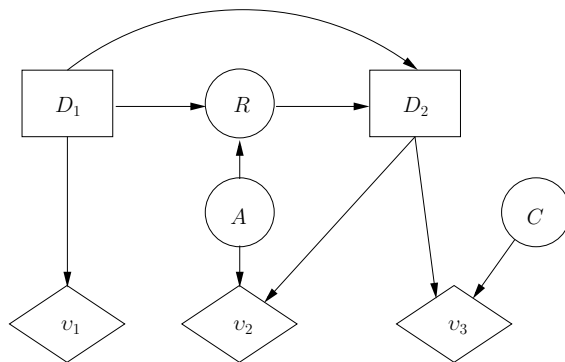


Figure 1: An influence diagram

IDs were initially proposed as a front-end for decision trees in the sense that their compact representation facilitated the problem structuring but they were later converted into a decision tree during the solution process. Olmsted (1983) described a method to solve IDs directly without converting them into decision trees. Shachter (1986) published the first ID evaluation algorithm. Evaluating an ID means computing a strategy that has the maximum expected utility. Rather than using arc reversals as in Olmsted’s and Shachter’s algorithms, other algorithms based on variable elimination strategies or on clique-trees approaches may be now used to solve IDs, Cooper (1988); Shenoy (1992); Jensen et al. (1994); Shachter and Ndilikilikisha (1993); Zhang (1998); Madsen and Jensen (1999). Although far from negligible, computational issues related to ID evaluation are beyond the scope of this paper. Some critical difficulties and their solutions are discussed and exemplified in Gómez et al. (2007); Bielza et al. (2000), where a large ID, called *IctNeo*, models neonatal jaundice management for an important public hospital in Madrid.

This paper focuses on modeling issues. IDs have an enormous potential as a model of uncertain knowledge. The process itself of building an ID provides a deeper understanding of the problem, and ID outputs are remarkably valuable. Given a specific configuration of variables, an ID yields the *best course of action*. But ID responses are not limited to provide optimal strategies of the decision-making

problem. *Inferred posterior distributions* may be employed to generate diagnosis outputs (probabilities of each cause). IDs may also automatically generate *explanations* of their proposals as a way to justify their reasoning, Fernández del Pozo et al. (2005); Bielza et al. (2008). Even worse, the domain expert may formulate a more difficult *query*, without specifying all the variables required to determine the optimal decision, leading to imprecise responses that should be refined if we want the decision maker to be satisfied, Fernández del Pozo and Bielza (2002). Reasoning in the reverse direction, assuming that the final results of the decisions are known, IDs can be used to generate *probabilistic profiles* that fit these final results (answering questions like “which kind of patients receive this specific treatment?”). Also, the computation of the *expected value of information* have shown to play a vital role in assessing the different sources of uncertainty, Shachter (1999).

The aforementioned special issue of *Decision Analysis* devoted to IDs is a sign of the lively interest in IDs. Boutilier (2005) discusses the profound impact that IDs have had on artificial intelligence. As a professional decision analyst, Buede (2005) reports on the value of IDs for tackling challenging real decision problems and considers IDs almost as indispensable as a laptop computer. Pearl (2005) recognizes the significant relevance of IDs but he underscores some limitations. First, due to their initial conception with emphasis on subjective assessment of parameters, econometricians and social scientists continued using traditional *path diagrams* where parameters were inferred from the data itself. Second, artificial intelligence researchers, with little interaction with decision analysis researchers at that time (early 1980s), established conditional independence semantics through the d-separation criterion developing competitive computational tools. Thus, although IDs are informal precursors to Bayesian networks, the former had a milder influence on automated reasoning research than the latter. Finally, Pauker and Wong (2005) consider that IDs have disseminated slowly in the medical literature, Owens et al. (1997); Nease and Owens (1997), compared to the dominating model of decision trees, the reasons remaining unclear.

This paper provides an outlook of ID modeling, discussing the recent contributions. It is our aim to contribute at getting closer the theoretical and practical developments of IDs fostering their wide use. Also, the analysis carried out identifies challenges and provides insights into the community involved in the design of decision models by means of IDs.

The paper is organized as follows. Section 2 discusses the representation of asymmetric decision problems. Section 3 discusses the sequencing of decisions and observations that, when partially specified (or unspecified), leads to another kind of asymmetry. Section 4 discusses the use of continuous decisions and uncertainties including the use of continuous conditionally deterministic variables. Section 5 discusses some challenges of representing decision problems in practice. Finally, section 6 finishes off with a summary, discussion, and an outline of other issues not discussed in this paper.

## 2 Asymmetric Decision Problems

In this section, we will define asymmetric decision problems and review various techniques that have been proposed in representing such problems based on influence diagrams.

In a decision tree representation, a path from the root node to a leaf node is called a *scenario*. A decision problem is said to be *symmetric* if (i) in all its decision tree representations, the number of scenarios is equal to the cardinality of the Cartesian product of the state spaces of all chance and decision variables, and (ii) there exists a decision tree representation of the problem such that the sequence of variables is the same in all scenarios. A decision problem is said to be *asymmetric* if it is not symmetric.

We will illustrate an asymmetric problem using the reactor problem, which was originally described by Covaliu and Oliver (1995), and subsequently modified by Bielza and Shenoy (1999). An electric utility firm has to decide whether to build ( $D_2$ ) a reactor of advanced design ( $a$ ), conventional design ( $c$ ), or no reactor ( $n$ ). If the reactor is successful, i.e., there are no accidents, an advanced reactor is more profitable, but it is riskier. Experience indicates that a conventional reactor ( $C$ ) has probability 0.98 of being successful ( $cs$ ) and 0.02 of being a failure ( $cf$ ). On the other hand, an advanced reactor ( $A$ ) has probability 0.66 of being successful ( $as$ ), probability 0.244 of a limited accident ( $al$ ), and probability 0.096 of a major accident ( $am$ ). If the firm builds a conventional reactor, the profits are estimated to be \$8B if it is a success, and  $-\$4B$  if it is a failure. If the firm builds an advanced reactor, the profits are \$12B if it is a success,  $-\$6B$  if there is a limited accident, and  $-\$10B$  if there is a major accident. The firm's utility function is linear in dollars. Before making a decision to build, the firm has an option to conduct a test ( $D_1 = t$ ) or not ( $nt$ ) of the components of the advanced reactor at a cost of \$1B. The test results ( $R$ ) can be classified as bad ( $b$ ), good ( $g$ ), or excellent ( $e$ ). Figure 2 shows a causal probability model for  $A$  and  $R$ . Notice that in the probability model, if  $A = as$ , then  $R$  cannot assume the state  $b$ . Thus, if the test results are observed to be bad, then an advanced reactor will result in either a limited or major accident, and consequently, the Nuclear Regulatory Commission will not license an advanced reactor.

Figure 3 shows a decision tree representation of the reactor problem. The computation of the marginal for  $R$ , and the conditional for  $A$  given  $R$ , have been done using the causal probability model for  $A$  and  $R$ , and this computation is shown in Figure 4. In this figure, the leaves of the tree contain the joint probabilities for  $A$  and  $R$ . The optimal strategy is to do the test; build a conventional reactor if the test results are bad or good, and build an advanced reactor if the test results are excellent. The expected profit associated with this strategy is \$8.130B.

The decision tree representation, shown in Figure 3, easily captures the asymmetric structure of the reactor problem. The cardinality of the Cartesian product of the state spaces of all variables is 108,

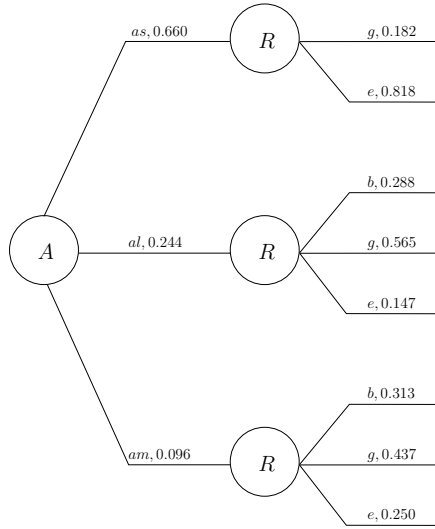


Figure 2: A probability model for  $A$  and  $R$  in the reactor problem



Figure 3: A decision tree representation of the reactor problem

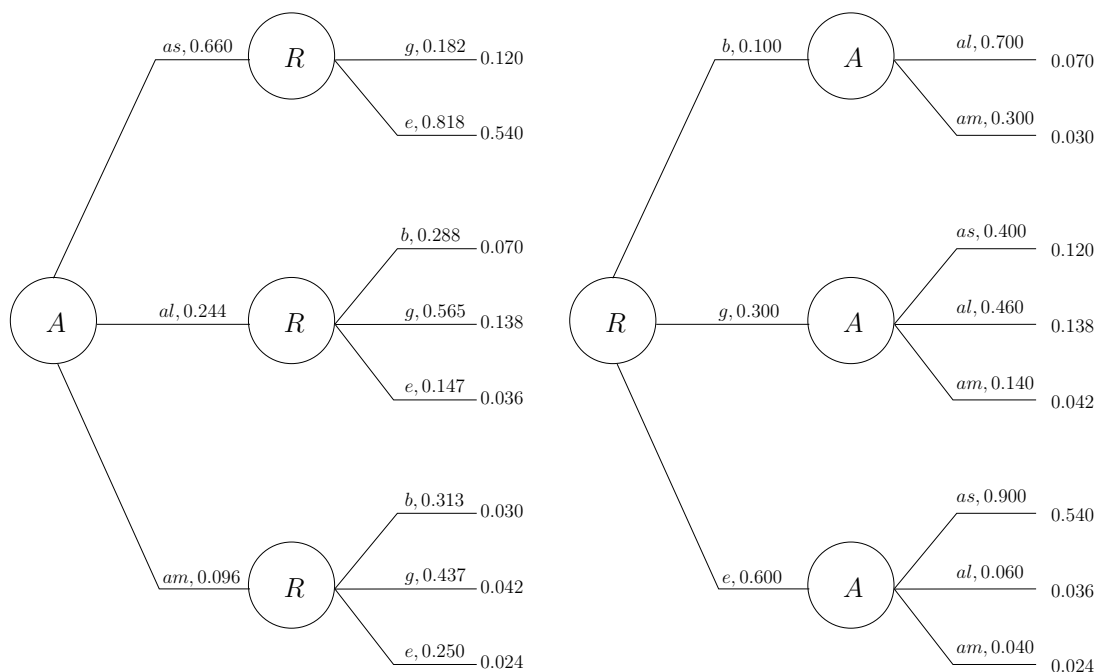


Figure 4: The pre-processing of probabilities in the reactor problem

but there are only 21 possible scenarios. The decision tree is shown using *coalescence*, i.e., repeating subtrees are shown only once.

There may be three kinds of asymmetry in asymmetric decision problems: chance, decision, and information. The state space of a chance variable may depend on the scenario, and in the extreme, a chance variable may be non-existent. For example, in the reactor problem,  $C$  does not occur in scenarios that include  $D_2 = a$ . This is called *chance* asymmetry. Similarly, the state space of a decision variable may depend on a scenario, and in the extreme, may be non-existent. For example, the alternative  $D_2 = a$  is not available in some scenarios. This is called *decision* asymmetry. Finally, the sequence of variables may depend on the scenario. For example,  $R$  does not precede  $D_2$  in all scenarios. This is called *information* asymmetry.

Since decision trees depict all possible scenarios, they can capture asymmetry easily. However, the number of nodes in a decision tree grows combinatorially with the number of variables, and are not tractable for large problems. Also, pre-processing of probabilities may be required (as in the reactor problem).

Influence diagrams were proposed by Howard and Matheson (1984) to avoid the problem of pre-processing in decision trees and to represent large problems that would be intractable using decision trees. However, IDs cannot capture asymmetric features of decision problems as easily as decision trees. One solution is to artificially enlarge the state spaces of chance and decision variables and embed the asymmetric problem in a higher dimensional symmetric problem. For example, an influence diagram

representation of the reactor problem is shown in Figure 5.

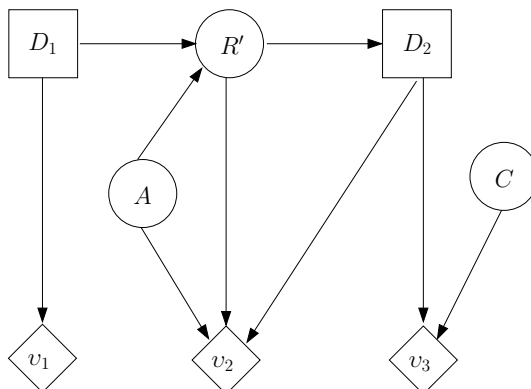


Figure 5: A symmetric ID representation of the reactor problem

Variable  $R$  in the decision tree has been replaced by  $R'$  which has states  $\{b, g, e, nr\}$ , and its conditional distribution now depends also on  $D_1$  as shown in Table 1. Thus, if  $D_1 = nt$ , then  $R' = nr$  with probability 1, and if  $D_1 = t$ , then we have the conditional probability distribution as before. Arc  $(R', D_2)$  means the true value of  $R'$  is known when an alternative of  $D_2$  has to be chosen.  $v_1$ ,  $v_2$  and  $v_3$  are additive factors of the joint utility function. The constraint that the alternative  $D_2 = a$  is not available when  $R' = b$  is modeled by making  $R'$  a parent of  $v_2$  and making the value of  $v_2(b, a, as) = v_2(b, a, al) = v_2(b, a, am) = -M$ , where  $M$  is a large positive number. This will ensure that when  $R' = b$ ,  $D_2 = a$  will never be optimal, and the constraint will be satisfied.

$P(R' D_1, A)$	$(nt, as)$	$(nt, al)$	$(nt, am)$	$(t, as)$	$(t, al)$	$(t, am)$
$b$	0	0	0	0	0.288	0.313
$g$	0	0	0	0.182	0.565	0.437
$e$	0	0	0	0.818	0.147	0.250
$nr$	1	1	1	0	0	0

Table 1: The conditional probability distribution for  $R'$  given  $D_1$  and  $A$

This completes a symmetric influence diagram representation of the reactor problem. Notice that we have increased the size of the problem in the sense that if we convert the influence diagram into a decision tree, it would result in 104 scenarios, whereas the asymmetric decision tree (shown in Figure 3) only has 21 scenarios. Thus, the symmetric influence diagram representation is not very efficient.

To address the inefficiency of the symmetric ID technique, Smith et al. (1993) propose representing the asymmetric aspects of a decision problem in the conditional distributions, leaving the overall ID graphical representation unchanged. The conditional distributions are represented using graphical structures called *conditional distribution trees* consisting of conditioning scenarios and atomic distributions. For example, in the reactor problem, the conditional distribution trees for  $D_2$  and  $R'$  are shown

in Figure 6.

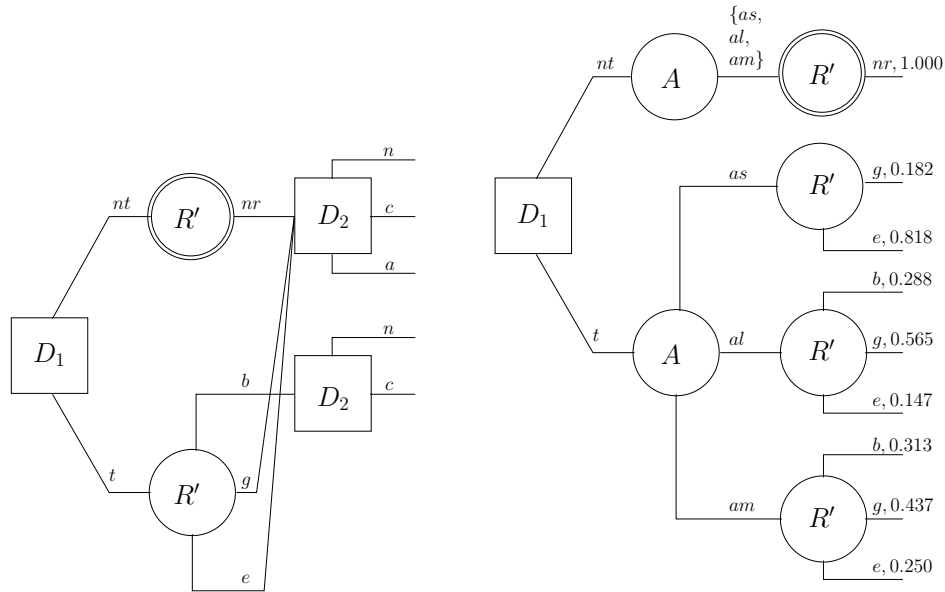


Figure 6: The conditional distribution trees for  $D_2$  and  $R'$

One advantage of conditioning trees is that the number of scenarios is trimmed back to the minimum necessary. Also, constraints on decision variables can be represented directly without having to add large negative utilities. Thus, we do not need to include an arc from  $R'$  to  $v_2$  in the ID graph. However, some disadvantages remain, such as the artificial state  $nr$  for  $R'$ , having  $D_1$  as a parent of  $R'$ , an arc  $(D_1, D_2)$  needs to be added to the ID graph, etc. Bielza and Shenoy (1999) discuss other strengths, weaknesses and open issues associated with the conditional distribution tree technique.

Next, we will discuss another approach to representing asymmetric decision problems called *sequential decision diagrams* proposed by Covaliu and Oliver (1995). Unlike the conditioning tree technique, sequential decision diagrams (SDDs) represent the asymmetric structure of the decision problem directly in the SDD graph. The SDD graph for the reactor problem is shown in Figure 7.

Like in IDs, each variable appears once in a SDD. The arcs in a SDD denote the sequencing of variables in scenarios like in decision trees. Arcs may be annotated by conditions and constraints. For example, arc  $(D_1, D_2)$  has a condition  $D_1 = nt$ , which means that whenever  $D_1 = nt$ , the next variable in the scenario is  $D_2$ . If there are any constraints on decisions, then this can also be listed on the corresponding arc. For example,  $D_2 = a$  is only possible if  $D_1 = nt$ , or if  $D_1 = t$  and  $R \neq b$ . The details of the probability model and utility function are not depicted in the SDD representation. Covaliu and Oliver assume that these details are represented by a corresponding influence diagram as in Figure 5.

SDDs can represent the set of all scenarios compactly. Also, we do not need to add dummy states to variables (such as  $R = nr$  as ID representations require). On the other hand, there are some drawbacks.

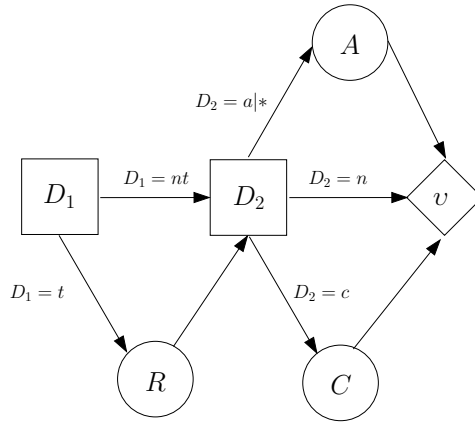


Figure 7: A sequential decision diagram for the reactor problem. \* denotes the constraint  $(D_1 = nt) \vee (D_1 = t \wedge R \neq b)$

First, SDDs cannot represent probability models consistently. For example, the ID representation has state space  $\Omega_R = \{nr, b, g, e\}$  for  $R'$ , whereas the SDD representation has  $\Omega_R = \{b, g, e\}$  for  $R$ . Second, SDD representation assumes that we have an unfactored joint utility function. This implies that solving a SDD cannot be done as efficiently (as we could have if we had an additive factorization of the joint utility function).

Demirer and Shenoy (2006) propose a representation called *sequential valuation networks* (SVNs) that is a hybrid of Shenoy's (2000) asymmetric valuation network and Covaliu and Oliver's SDDs. A SVN representation of the reactor problem is shown in Figure 8.

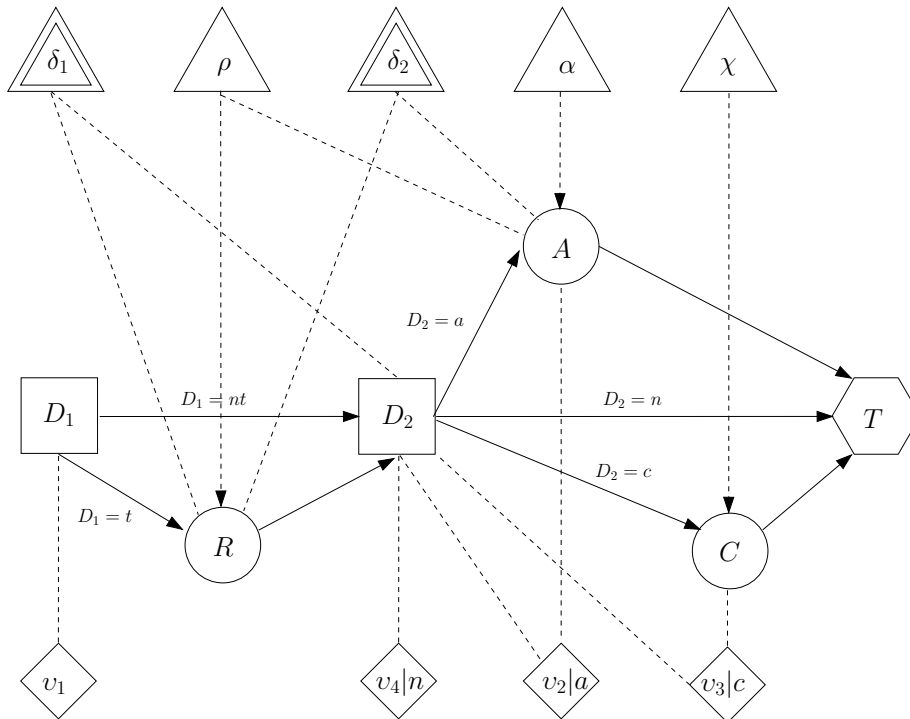


Figure 8: A sequential valuation network for the reactor problem

The SVN consists of an SDD plus indicator, probability and utility valuations. The SVN has an hexagonal terminal node (instead of the diamond node in the SDD) whose only role is to signify the end of a scenario. Probability valuations are shown as single-bordered triangular nodes. The dashed edges connecting probability valuations to variables denote the domain of the probability valuation. Thus, the domain of  $\rho$  is  $\{A, R\}$ , the domain of  $\alpha$  is  $\{A\}$ , and the domain of  $\chi$  is  $\{C\}$ . If the probability valuation is a conditional for a subset of its domain given the rest of the domain, then the dashed edges to the variables in the subset are directed towards them. Thus,  $\rho$  is a conditional for  $\{R\}$  given  $\{A\}$ ,  $\alpha$  is a conditional for  $\{A\}$  given  $\emptyset$ , and  $\chi$  is a conditional for  $\{C\}$  given  $\emptyset$ . The set of all probability valuations denotes a multiplicative factorization of the joint probability distribution of the chance nodes in the SVN representation. Indicator valuations encode qualitative constraints on the joint state spaces of variables and are represented as double-bordered triangular nodes in the SVN graph. The reactor problem representation has two indicator valuations.  $\delta_1$  with domain  $\{R, D_2\}$  removes the state  $(b, a)$  from the joint state space of  $\{R, D_2\}$ , and  $\delta_2$  with domain  $\{R, A\}$  removes the state  $(b, as)$  from the joint state space of  $\{R, A\}$ . Utility valuations represent additive factors of the joint utility function and are represented by diamond shaped nodes, and the dashed edges between the utility valuations and variables denote the domain of the valuations. In practice, many of the values of the utility factors are zeroes and need not be specified. For example, the costs associated with building an advanced reactor are only associated with  $D_2 = a$ . Thus, this factor is defined on the state space  $\{a\} \times \Omega_A$ . This factor is denoted in the SVN representation by  $v_2|a$ . Similarly,  $v_3|c$  with domain  $\{D_2, C\}$ , and  $v_4|n$  with domain  $\{D_2\}$  are utility valuations defined on  $\{c\} \times \Omega_C$  and  $\{n\}$ , respectively.

Besides the representations already discussed, there are a number of other graphical representations proposed in the literature for asymmetric decision problems. These include Call and Miller (1990) combination of symmetric influence diagrams and decision trees, Qi et al. (1994) decision graphs, contingent IDs (Fung and Shachter (1990)), Shenoy (2000) asymmetric valuation networks, Nielsen and Jensen (2003) asymmetric influence diagrams, Jensen and Vomlelova (2002) unconstrained influence diagrams, Liu and Shenoy (2004) coarse valuation networks, and Jensen et al. (2006) sequential influence diagrams.

In spite of the plethora of different graphical representations, some aspects of asymmetric decision problems are difficult to represent. One example is a certain event that is repeated for a random number of times. Thus, e.g., we can often find medical situations where a specific treatment cannot be performed more than e.g. twice per full treatment of a patient due to its risks. Also, that risky treatment must be followed and preceded by a certain control treatment, etc. Traditional IDs could not meet these constraints, containing sequences of treatments that are impossible. For another example, a physician has to decide on a subset of possible tests and on a sequence for these tests until a

satisfactory diagnosis can be made. If the set of possible tests is large, the combinatorial explosion of the sequencing of all subsets makes the problem intractable for representation and solution. We will discuss this problem in more detail in the next section.

In summary, IDs provide a compact graphical method for representing and solving symmetric decision problems. For asymmetric decision problems, several alternative representations have been proposed, including conditional distribution trees, sequential decision diagrams, and sequential valuation networks. While there have been some studies (e.g., Bielza and Shenoy (1999)) that compare some of the representations, more studies are needed to compare all of the different proposed representations.

### 3 Sequencing of Decisions and Uncertainties

In this section, we will deal with information constraints that specify the uncertainties/decisions that are observed and those that are unobservable at the time a decision has to be made. It also includes the relaxation of the total order of decisions—sequencing constraint—and even between decisions and uncertainties.

In a symmetric decision problem, the information constraints are the same in all scenarios. Thus, in the symmetric ID representation of the reactor problem shown in Figure 5, for decision  $D_1$  (whether to do a test of the advanced reactor or not), all uncertainties/decisions are unobservable, and for decision  $D_2$  (type of reactor to build), the results of the test  $R'$  and decision  $D_1$  are observable, but the true states of  $C$  and  $A$  are unobservable. This is represented in the symmetric ID shown in Figure 5 as follows. Since all uncertainties/decisions are unobservable for  $D_1$ , there are no arrows into  $D_1$ . Since the results of test  $R'$  are observable for  $D_2$ , we have  $(R', D_2)$ . Since we have a directed path from  $D_1$  to  $D_2$ , and we are assuming the no-forgetting condition, the decision made at  $D_1$  is known when a decision at  $D_2$  has to be made. This can be modeled explicitly by including an arc  $(D_1, D_2)$ , or implicitly by not having the arc, but assuming the no-forgetting condition during the evaluation process.

In asymmetric decision problems, the information constraints may not be the same in all scenarios. For example, in the decision tree representation of the reactor problem (shown in Figure 3),  $R$  is an information predecessor of  $D_2$  in some scenarios (when  $D_1 = t$ ), but not in some others (when  $D_1 = nt$ ). While this problem can be made symmetric as described in the previous section, the technique described there—adding artificial states to some chance variables—cannot be applied in all problems. This is especially true if the chance variable being modified is a parent of other chance variables (since if we add an artificial state to such a variable, then we have to make up conditional distributions for each of the child variables that is consistent with the original joint distribution, which is a non-trivial task).

Consider a problem of diagnosing diabetes. Diabetes can be detected by doing a blood test and measuring the level of Glucose, which is elevated for diabetic patients. A less expensive test is to take a urine specimen and measure the level of Glucose in urine. A physician is trying to decide whether or not to treat a patient for diabetes. Before she makes this decision, she can order a test (blood or urine). Based on the results of the first test, she can then decide whether or not to order a second test. A symmetric ID model for this problem is shown in Figure 9.

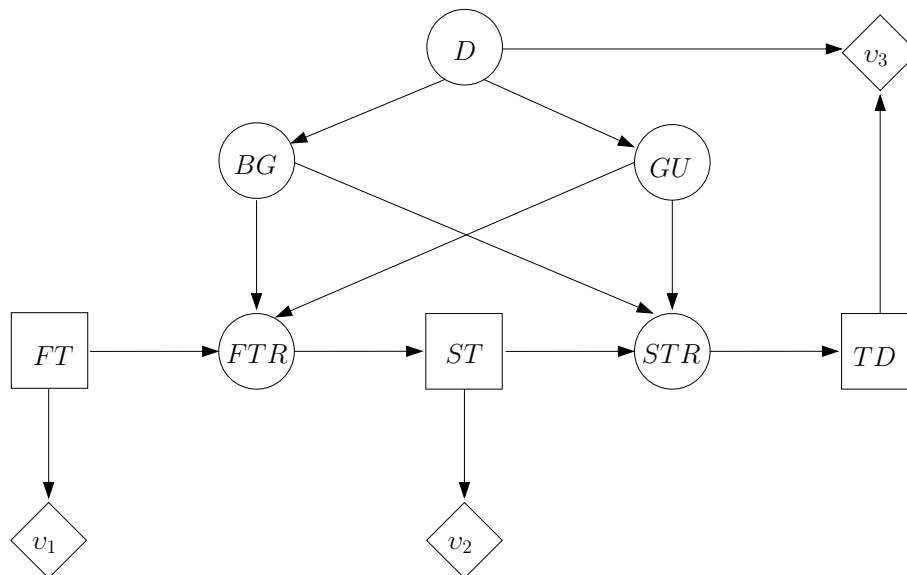


Figure 9: A symmetric ID model for the diabetes diagnosis problem

In this model,  $D$  is the diabetes variable with states  $d$  (has diabetes) and  $nd$  (no diabetes),  $BG$  is the blood Glucose chance variable with states  $bge$  (elevated) and  $bgn$  (normal),  $GU$  is the Glucose in urine chance variable with states  $gue$  (elevated) and  $gun$  (normal),  $FT$  is the first test decision variable with states  $bt$  (blood test),  $ut$  (urine test), and  $nt$  (no test),  $FTR$  is the first test results chance variable with states  $bge$ ,  $bgn$ ,  $gue$ ,  $gun$ , and  $nr$  (no results),  $ST$  is the second test decision variable with the same states as  $FT$ ,  $STR$  is the second test results chance variable with the same states as  $FTR$ , and  $TD$  is the treat for diabetes decision variable with states  $td$  (treat for diabetes) and  $ntd$  (not treat for diabetes).  $v_1$ ,  $v_2$ , and  $v_3$  are additive factors of the joint utility function.  $FTR$  and  $STR$  are conditionally deterministic variables with conditional distributions as follows. If  $FT = bt$ , and  $BG = bge$ , then  $FTR = bge$  with probability 1, etc. This formulation allows the possibility of repeating the same test and if the costs of the tests are positive, it is never optimal. An advantage of this formulation is that we leave the original Bayesian network consisting of arcs  $(D, BG)$ , and  $(D, GU)$  unchanged with no artificial states added to the chance variables. A disadvantage of this formulation is that we add two artificial chance variables  $FTR$  and  $STR$  with large state spaces. Thus, if we have a decision problem with, say 20 possible tests, formulating a symmetric ID in this fashion would be

intractable to solve.

To deal with diagnosis problems with many possible tests and no ordering specified among them, Jensen and Vomlelova (2002) propose a representation called *unconstrained influence diagrams* (UIDs). In UIDs, the requirement that there exists a directed path that includes all decision variables is dropped. The order of tests is deliberately unspecified so that there is a partial order among the decision variables. In the solution phase, we need to determine the tests to be done in some sequence as a function of what is observed from the results of the tests that are already done. A UID model for the diabetes diagnosis problem is shown in Figure 10.

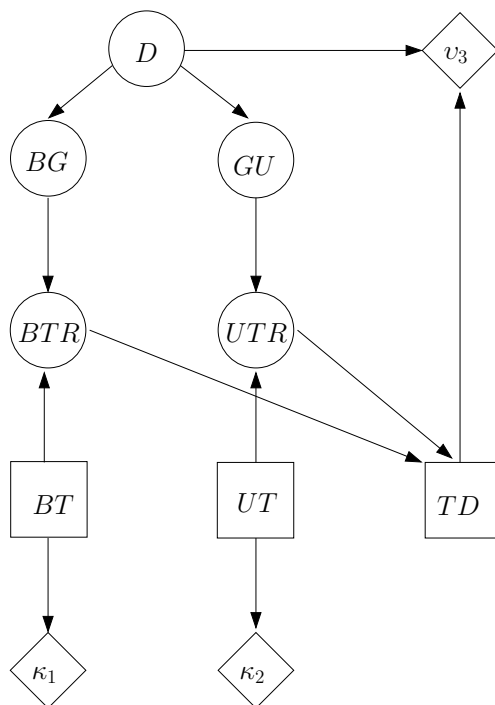


Figure 10: An unconstrained ID model for the diabetes diagnosis problem

In this model,  $D$ ,  $BG$ ,  $GU$ ,  $TD$  are as before.  $BT$  is a decision variable with states  $bt$  and  $nt$ , and  $UT$  is a decision variable with states  $ut$  and  $nt$ .  $BTR$  (for blood test results) is a chance variable with states  $bge$ ,  $bgn$ , and  $nr$ , and  $UTR$  (for urine test results) is a chance variable with states  $gue$ ,  $gun$ , and  $nr$ .  $BTR$  and  $UTR$  are conditionally deterministic variables such that if  $BT = bt$  and  $BG = bge$ , then  $BTR = bge$  with probability 1, etc. If  $BT = nt$ , then  $BTR = nr$  with probability 1 (regardless of the state of  $BG$ ). Similarly for  $UTR$ . Notice that there is only a partial order among the decision variables.  $\{BT, UT\}$  precede  $TD$ . The ordering between  $BT$  and  $UT$  is left unspecified. An optimal solution would include a sequencing of some tests as a function of what is observed from the previous tests performed. If  $BT$  is the first test and  $UT$  is the second test, then  $BTR$  becomes an informational predecessor of  $UT$ , etc.

An advantage of UIDs is that the state spaces of some of the decision and chance variables are

much smaller than in a corresponding symmetric ID. For example, the state spaces of  $BT$  and  $UT$  have cardinality two, whereas in the symmetric ID representation,  $FT$  and  $ST$  have cardinality three. Also, the state spaces of  $BTR$  and  $UTR$  have cardinality three, whereas the corresponding variables  $FTR$  and  $STR$  have cardinality five. Furthermore,  $BTR$  and  $UTR$  each have two parents, whereas the corresponding variables  $FTR$  and  $STR$  each have three parents. (We should note here that  $BTR$  and  $UTR$  have different semantics than  $FTR$  and  $STR$ . These are synthetic variables that are constructed to represent the information asymmetry inherent in the decision problem.) These differences in the number of states are even more accentuated with the number of tests. Thus, if we have a diagnosis problem with say 10 possible tests, a UID representation may be tractable for solution even though a symmetric ID is not.

UIDs are designed to simplify the representation of diagnosis/troubleshooting problems only. The evaluation procedure for solving UIDs is much more complex than the procedure for solving a symmetric ID since the sequencing of the tests is not specified and has to be determined in the solution phase. When the number of possible tests is large, solving a UID may also be intractable.

In principle, a decision tree representation should be able to easily represent the asymmetric features of the diabetes diagnosis problem. A drawback would be the size of the decision tree. Since sequential valuation networks are able to represent a decision tree compactly, this representation could be used to represent this problem using no artificial variables. A SVN representation of the diabetes diagnosis problem is shown in Figure 11. In this model,  $FT$  and  $ST$  are decision variables with states  $bt$ ,  $ut$ , and  $nt$ . The chance variables  $D$ ,  $BG$ , and  $GU$  have no dummy states. The model allows repetition of tests, but assuming the tests have positive costs, this is never optimal.

The diabetes diagnosis problem can also be represented by a *sequential influence diagram*, Jensen et al. (2006). A sequential influence diagram (SID) can be considered as a combination of sequential decision diagrams, influence diagrams, and unconstrained influence diagrams. There are two types of arcs in a SID—sequential and dependence. To distinguish between the two, we show sequential arcs by solid lines and dependence arcs by dotted lines. If we have a partial order on the information constraints, we use clusters and sequential arcs between clusters to represent partial orders. A SID representation of the diabetes diagnosis problem is shown in Figure 12.

In this model, arcs  $(D, BG)$  and  $(D, GU)$  constitute a Bayesian network without any dummy states for  $BG$  and  $GU$ .  $BT$  and  $UT$  are decision nodes with state spaces  $\{bt, nt\}$ , and  $\{ut, nt\}$ , respectively. The sequential arc  $(BT, BG)$  with the annotation  $BT = bt$  denotes that  $BG$  follows  $BT$  only in scenarios where  $BT = bt$ . Thus, if  $BT = nt$ , then  $BG$  does not follow  $BT$ . Similarly for the sequential arc  $(UT, GU)$  with the annotation  $UT = ut$ . The cluster of nodes containing  $BT$ ,  $UT$ ,  $BG$  and  $GU$  represents the fact that the sequence of tests is unspecified and to be determined.

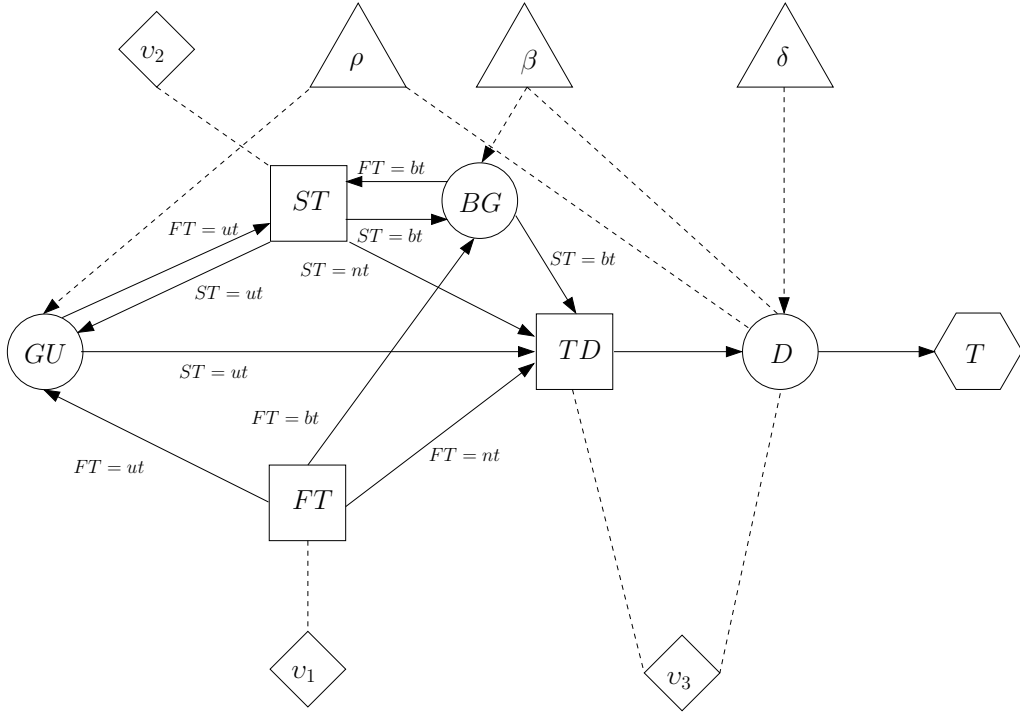


Figure 11: A sequential valuation network model for the diabetes diagnosis problem

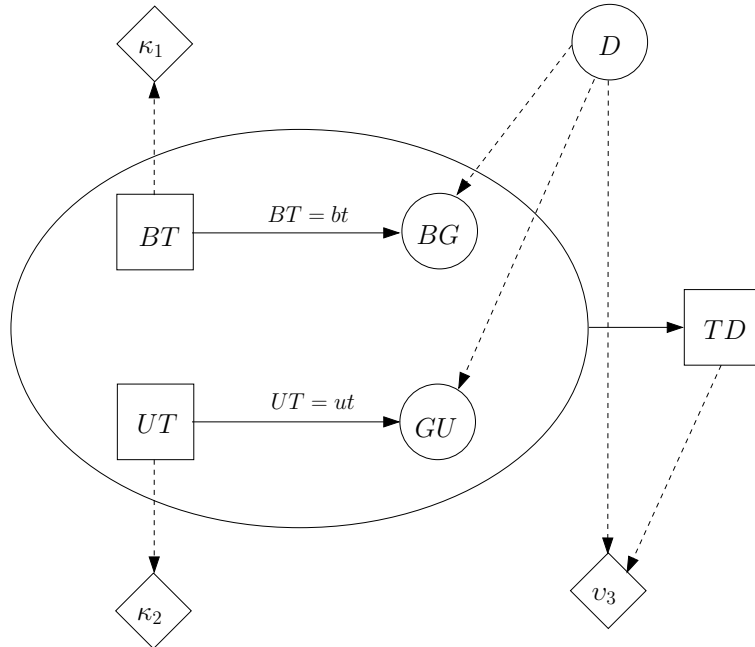


Figure 12: A sequential ID model for the diabetes diagnosis problem

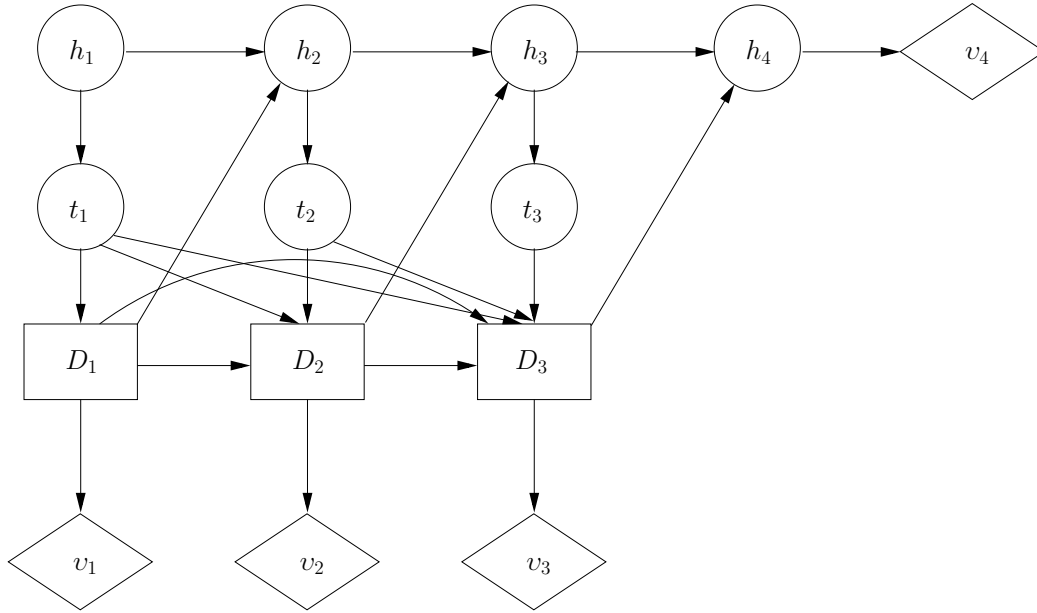
However, the sequential arc from this cluster to  $TD$  specifies that in all scenarios,  $TD$  follows the two tests. Finally,  $D$  follows  $TD$  in all scenarios. The remaining part of the diagram has the same semantics as an influence diagram. If we compare this model with the UID model in Figure 10, we notice that we do not need dummy variables  $BTR$  and  $UTR$ . If we compare the SID model with the SVN model shown in Figure 11, we notice that the sequence of tests is represented explicitly in the SVN representation, while it is represented implicitly in the SID representation. Thus, if we have 10 different tests, a SVN representation can get complex to account for all possible sequences of tests, whereas the SID representation remains simple. The SID representation is more efficient than either the UID representation or the SVN representation.

We conclude this section with a discussion of partial influence diagrams (PIDs) and limited memory influence diagrams (LIMIDs). The algorithms for solving IDs assume that there is a total ordering of the decision variables. This condition is sufficient for computing an optimal strategy, but it is not necessary. Nielsen and Jensen (1999) define *partial IDs* (PIDs) as IDs where there is no total order specified for decision variables. Since the solution to a decision problem may depend on the ordering of the decision variables, Nielsen and Jensen specify conditions that ensure that a PID has a well-defined optimal strategy even though there is no total order on the decision variables. The conditions are based on d-separation and can be read from the graphical structure of a PID.

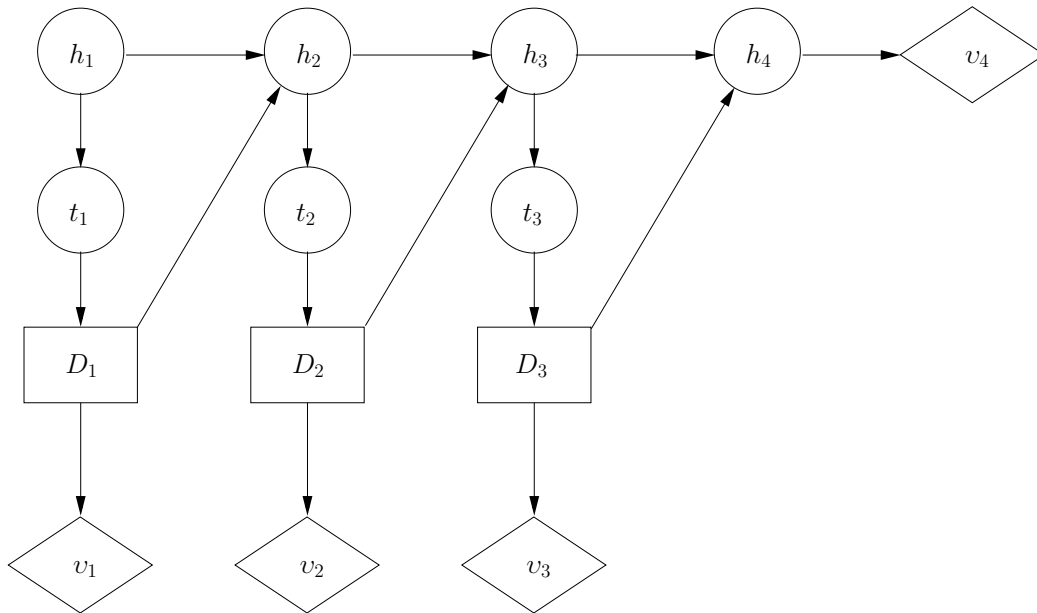
Lauritzen and Nilsson (2001) introduce the notion of *limited memory* IDs (LIMIDs) as a model for multistage decision problems in which two conditions are relaxed: total order for decision variables and no-forgetting. The sequence in which decisions are to be made is not specified other than through it being compatible with the partial order induced by the ID graph, i.e., if  $D_2$  is a descendant of  $D_1$ , then decision  $D_1$  must be made before  $D_2$ . The parents of a decision node  $D$  represent exactly the variables whose values are known and taken into consideration when a decision at  $D$  has to be made. Thus, LIMIDs allow for multiple agents of a decision maker (who may not be able to communicate with each other) or for forgetful decision makers.

The example of LIMID presented in Lauritzen and Nilsson (2001) describes a pig breeder problem involving decisions about injecting a certain drug as a treatment for a given disease. The disease can be detected with a test ( $t_i$ ). The decisions about injections must be done during the first three months of pigs life, one per month ( $D_i$ ,  $i=1,2,3$ ). The pigs will be sold in the fourth month. The market price will depend on the health status of the pigs ( $h_i$ ). This problem can be considered with two different points of view. The first one assumes that the breeder keeps individual records for every pig, and this information is used before deciding about the injections. Thus, the no forgetting condition is assumed, see Figure 13(a). The second assumes that there are no individual records for the pigs, and that the decisions are made knowing the test result only for the given month (the LIMID version, see

Figure 13(b)). Nodes  $v_i, i = 1, 2, 3$  are related to the costs of performing the test and  $v_4$  denotes the benefit obtained with the pig sale.



(a) ID version



(b) LIMID version

Figure 13: The ID and limited memory ID model for the pig problem

The main motivation of LIMIDs is that its solution is computationally tractable whereas the same decision problem could be intractable if we assumed the no-forgetting condition. The disadvantage is that an optimal solution of a LIMID may not be optimal for the same decision problem if we were to assume the no-forgetting condition. Thus, we trade-off tractability of solution with optimality. Lauritzen and Nilsson describe an algorithm called *single policy updating* for finding a local optimal

solution of a LIMID.

In summary, we have examined different representations of problems in which we have information asymmetry. We have examined symmetric influence diagrams, unconstrained influence diagrams, sequential valuation networks, sequential influence diagrams, partial influence diagrams, and limited memory influence diagrams. For the class of problems in which we have several diagnostic tests that can be done in any sequence, the sequential influence diagram representation is very efficient in representing such problems. The exact solution of these problems, however, is hard and may not be tractable when we have a large number of tests. In some decision problems, there may not be a total order specified among the decision nodes, but it admits a well-defined optimal solution through the use of partial influence diagrams. Also, for a class of multi-stage problems, influence diagrams may be intractable. By dropping the no-forgetting condition, the solution of a decision problem may be tractable as in limited memory influence diagrams.

## 4 Continuous Chance and Decision Variables

In this section, we review the literature on the use and issues related to using continuous chance and decision variables in representing and solving decision problems.

In practice, it is safe to assume that one encounters chance and decision variables that are continuous. A chance or decision variable is said to be discrete if the state space is countable, and continuous otherwise. Typically, the states of discrete variables are symbols, whereas the states of continuous variables are real numbers. The conditional distribution of continuous chance variables are conditional probability density functions (pdfs). One major problem associated with solving IDs with continuous chance variables is integration of products of conditional pdfs and continuous utility functions when computing expected utilities. There are many commonly used pdf that cannot be integrated in closed form. Another major problem is finding the maximum of a utility function on the state space of a continuous decision variable. Depending on the nature of the multidimensional utility function, we can have a difficult non-linear optimization problem. To avoid these difficulties, one standard approach is to discretize the continuous variables using bins. Using many bins results in a computational burden for solving the discrete ID, and using too few bins leads to an unacceptable approximation of the problem.

One of the earliest work on using continuous variables in IDs is Kenley (1986); Shachter and Kenley (1989). In their representation, which they call *Gaussian IDs* (GIDs), all chance variables are continuous having the so-called conditional linear Gaussian (CLG) distribution. These are Gaussian distributions whose mean is a linear function of its parents, and the variance is a constant. An implication of this condition is that the joint distribution of all chance variables is multivariate normal.

One can find marginals of the multivariate normal distribution without doing any integration. So the problem of integration does not exist for such IDs. GIDs also assume that all decision nodes are continuous and that the utility function is a quadratic function of its parent chance and decision variables. An implication of this condition is that there is a unique maximum for the decision variables, which can be found in closed form. Thus, the optimization problem of finding an optimal solution does not exist either for such IDs.

An example of a GID is shown in Figure 14 (Shachter and Kenley (1989)). A consultant has purchased an expensive computer for her practice. She will bill her clients for usage at an hourly rate, but expects to have spare capacity, which she could sell to time-share users. She needs to decide on a price for her clients and time-share users that will maximize her expected profit.

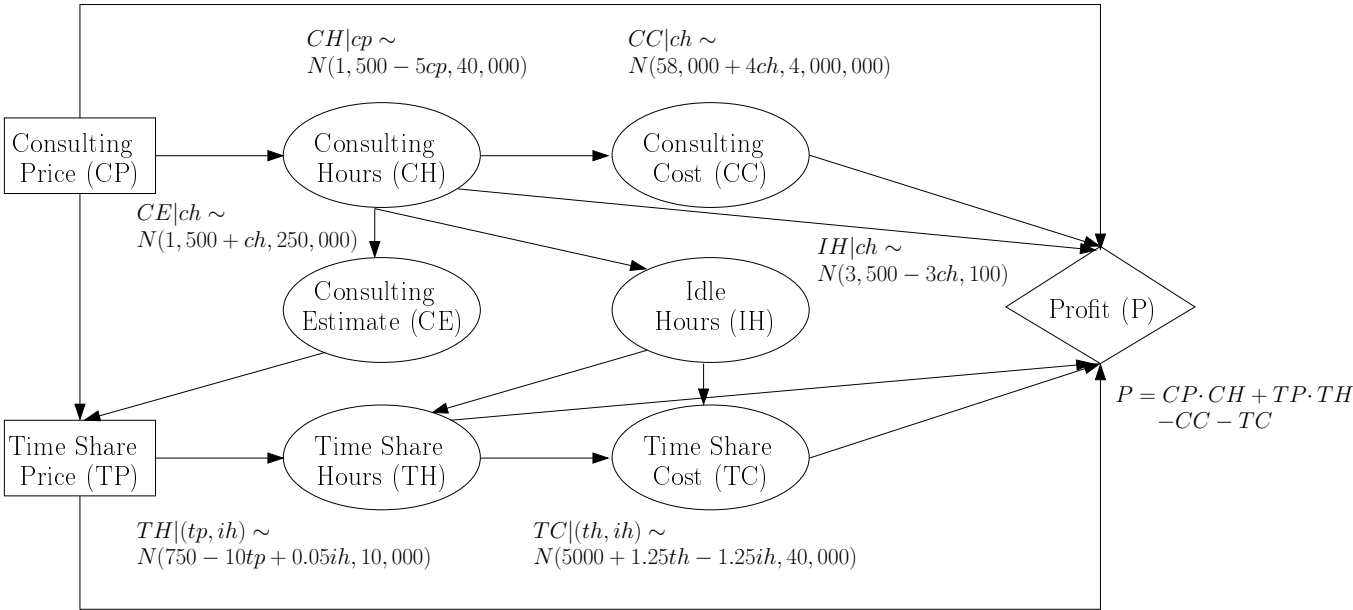


Figure 14: A Gaussian ID for the Consultant's Problem

Poland III (1994) extends GIDs to mixture of Gaussians IDs (MoGIDs). In MoGIDs, we have discrete and continuous chance variables. As in GIDs, continuous variables have a conditional linear Gaussian distribution whose mean is a linear function of its continuous parents, and a constant variance. The linear function and constant variance can be different for each state of its discrete parents. However, discrete nodes cannot have continuous parents. An implication of these conditions is that given a state of each discrete variable, the joint distribution of the continuous variables is multivariate normal. Again the problem of integration does not exist. MoGIDs have the same restriction as GIDs regarding the decision variables (all are continuous) and the utility function (quadratic).

If we have a continuous variable whose conditional pdf is not conditional linear Gaussian, then we can approximate it with a mixture of Gaussians. The parameters of the mixture distribution

(number of distributions, weights, and means and variances of the Gaussian distributions) can be found either by using a minimum entropy technique (Poland III and Shachter (1993)) or by solving a non-linear optimization problem that minimizes some distance measure between the target distribution and mixture distribution (Shenoy (2006)).

If we have a discrete chance variable with a continuous parent, then we do not have a mixture of multivariate normal distribution for the continuous chance variables. Some authors suggest approximating the product of the conditionals for the discrete variable and its parents by a CLG using a variational approach (Murphy (1999)) or a numerical integration called Gaussian quadrature (Lerner et al. (2001)). Shenoy (2006) suggests using arc reversals to remedy this situation. However, arc reversals may destroy the CLG distribution of the parent, in which case, it has to be approximated by a mixture of Gaussians.

If besides continuous chance variables, we have a decision problem with a (few) continuous decision variables, Bielza et al. (1999) suggest the use of a Markov Chain Monte Carlo (simulation) method for finding an approximately optimal strategy as a remedy for the optimization and integration problem. They define an artificial distribution on the product space of chance and decision variables, such that sampling from this artificial augmented probability model is equivalent to solving the original decision problem. The approach can accommodate arbitrary probability models and utility functions.

Charnes and Shenoy (2004) investigate the problem of solving large IDs with continuous chance variables and discrete decision variables using a Monte Carlo method where only a small set of chance variables are sampled for each decision node. Using this technique, they solve a problem of valuation of a Bermudan put option with 30 discrete decision variables and continuous chance variables with non-CLG distributions.

Another strategy for easing the problem of integration is suggested by Moral et al. (2001). They suggest approximating conditional pdfs by mixtures of exponential functions whose exponent is a linear function of the state of the variable and its continuous parents. Each mixture component is restricted to a hypercube. One advantage of this approximation is that such mixtures of truncated exponentials (MTEs) are easy to integrate in closed form. Also, the family of MTE functions are closed under multiplication and integration, operations used in solving influence diagrams. Cobb et al. (2006) describe MTE approximations of commonly used pdfs using an optimization technique similar to the one described earlier for finding parameters of a mixture of Gaussian approximation. Cobb and Shenoy (2008) define MTE IDs, where we have continuous chance variables with MTE conditional pdfs, discrete decision variables, and utility functions that are also MTE functions. Such MTE IDs can be solved using the solution technique of discrete IDs with integration used to marginalize continuous chance variables.

In the same spirit as MTEs, Shenoy and West (2009b) suggest the use of mixture of polynomials (MOPs) to approximate conditional pdfs. Like MTE functions, MOP functions are easy to integrate, and the family of MOP functions is closed under multiplication and integration. Unlike MTEs, finding MOP approximations is easier for differentiable functions as one can use the Taylor series expansion to find a MOP approximation. Finding a MTE approximation for a multi-dimensional distribution (such as the conditional for a variable given another continuous variables as parents) can be difficult, whereas finding a MOP approximation is easier as it can be found using the multi-dimensional version of the Taylor series.

In an ID, a continuous chance variable is said to be *deterministic* if the variances of its conditional distributions (for each state of its parents) are all zeroes. An example of a deterministic variable is a continuous variable whose state is a deterministic function of its continuous parents, and the function may depend on the state of its discrete parents. An example of a deterministic variable is  $Profit = Revenue - Cost$ , where *Revenue* and *Cost* are continuous chance variables, and *Profit* is a deterministic variable with *Revenue* and *Cost* as parents. Deterministic variables pose a special problem since the joint density for all the chance variables does not exist. For GIDs and MoGIDs, deterministic variables are not a problem as long as the functions defining the deterministic variables are linear. The theory of multivariate normal distributions allows linear deterministic functions (Rao (1973); Lauritzen and Jensen (2001)). The class of MTE functions is also closed under transformations required by linear deterministic functions (Cobb and Shenoy (2005a)), but not for non-linear deterministic functions. For non-linear deterministic functions, Cobb and Shenoy (2005b) suggest approximating a non-linear deterministic function by a piecewise linear function, and then using the technique proposed in Cobb and Shenoy (2005a). Cincicoglu and Shenoy (2009) describe an arc reversal theory for hybrid Bayesian networks with deterministic variables with a differentiable deterministic function. They use Dirac delta functions (Dirac (1927)) to represent deterministic functions. The deterministic function does not have to be linear or even invertible. As long it is differentiable and its real zeroes can be found, arc reversal can be described in closed form. They conjecture that Olmsted's arc-reversal theory (Olmsted (1983)) for solving discrete IDs can be used for hybrid IDs using their arc-reversal theory. This claim needs further investigation. The family of MOP functions are closed for a larger class of deterministic functions than MTEs. For example, MOP functions can be used for quotients (Shenoy and West (2009b)). The use of MOP functions for solving IDs with discrete, continuous, and deterministic variables needs to be investigated. Finally, deterministic variables pose no problems for Monte Carlo methods, such as Charnes and Shenoy (2004), that use iid samples. However, Markov Chain Monte Carlo methods do not converge in the presence of deterministic variables.

## 5 Modeling with IDs: Strengths and Limitations

The previous sections have presented ID modeling capabilities and challenges mostly developed and identified by researchers and published in scientific forums. But there is not too much feedback from analysts and experts about their experiences using IDs for building decision-making models. Specifically, it would be interesting to know what are the main problems perceived by those which have tried to model complex decision-making problems using IDs. Unfortunately, it is well known that the modeling step is not yet automated, and is considered as an art, and in which the most of the literature has taken little interest, Clemen (1997). This is the reason why there are few papers offering this point of view. Two examples of them are Bielza et al. (2000); Gómez et al. (2007).

These papers describe the problems faced while modeling a neonatal jaundice management problem in the medical domain. This problem is present during the first few days of a baby's life, the first 72 hours after birth being the most critical. The first decision is whether to admit (or not) a baby to a hospital and confining it, eventually, to an intensive care unit. In case the baby is admitted, it is necessary to control the bilirubin levels, carrying out different tests, and applying some of the prescribed treatments: phototherapy, exchange transfusion or observation. The treatment will be selected depending on some crucial factors such as age, weight, bilirubin and hemoglobin levels. Treatments are given along several consecutive stages, observing after each one their effects on the baby and repeating the process as many times as necessary until the problem is solved (the infant is then discharged or (s)he receives a treatment related to another disease).

The two main problems identified when constructing an ID for this disease are related to time modeling and the existence of constraints between the treatments (both are closely related). Since experts consider that there must be at least 6 hours between two successive treatments (in order to be able to observe their effects), the initial approach considers an ID containing a sequence of 12 decision nodes (one per stage). As treatments decisions depend at each stage on the same set of factors the whole model can be considered as a repetition of 12 identical slices, the  $i$ th-slice inheriting the information from  $(i - 1)$ th-slice. All decision nodes would be identical containing the same set of alternatives. Moreover, during the modeling stage, several constraints were identified: not to perform more than two exchanges per full treatment, start with observation or phototherapy, precede and follow exchanges by phototherapies, etc. These constraints lead to a highly asymmetric ID and it cannot be modeled with this structure of 12 identical decisions. Another problem related to time management stems from the variable length of the full process. Some patients will need only one treatment stage, for example. For that reason the subsequent decision domains must be filled in with dummy therapeutic actions (do nothing) in the spirit of symmetric IDs. Although direct, this first approximation entails a highly asymmetric and intractable ID due to the considerable set of nodes and arcs.

A deeper analysis revealed that the experts normally consider combinations of therapies: phototherapy of 12 or 18 hours long, for example. Using this fact, and in order to reduce the number of decisions, three types of frames are considered. The first one corresponds to the first decision, containing the alternatives allowed when starting the treatment. The second type will be used for two decision nodes articulating the central part of the treatment. Its alternatives consider grouped treatments satisfying the constraints (e.g. a 12-h phototherapy followed by an exchange transfusion and then a 12-h phototherapy again). The fourth and fifth decisions are related to the third type of frame. The fourth one concerns light treatments and the last only final actions. The consideration of combined actions makes more complex the definition of the decision domains, but reduces the model size to five decision nodes and limits the presence of incoherent treatments as well. There is another set of constraints not included in the new domains (just because they affect the whole process) and will have to be considered during the model evaluation.

Regardless of these difficulties, the modeling of this ID was described as a rich exercise for experts, who felt very comfortable reasoning about the different alternatives under consideration. This reasoning process led to a better understanding of the jaundice problem and, as a direct consequence, to a drastic reduction in the number of aggressive treatments. These positive outcomes were mostly a result of the modeling process, rather than of the model evaluation.

Another way to check whether the ID methodology is really useful in practice is to examine the literature on applications of decision problems using IDs. A compilation can be found in Gómez (2004). But this examination does not provide information about which systems are real applications and which are just prototypes. Real applications will refer to final systems in use and prototypes may be related to research projects without a commercial use.

More indicators of ID dissemination and acceptance can be obtained by looking for commercial software adding functionality for modeling and evaluating IDs. These tools would not exist if IDs were not considered useful. As a sample, we list seven relevant tools including the URL of the companies. Several of these URLs also contain links to case studies related to IDs, as well as a list of clients or partners who use IDs for facing their decision-making problems.

- **Analytica** (<http://www.lumina.com/>) is a visual tool for creating, analyzing, and communicating decision models as IDs.
- **Decision Manager** (<http://www.deciware.com>) offers tools for effective business management including decision trees and IDs.
- **DPL 7** (<http://www.syncopation.com>), a professional tool for decision analysis. It offers a single and coherent application with a straightforward graphical user interface.

- **HUGIN** (<http://www.hugin.com>) has focused on delivering advanced solutions for decision making under uncertainty in the financial service industry. It offers a complete and intuitive tool for building and analyzing ID models.
- **Netica** (<http://www.norsys.com/>) is a powerful and easy-to-use program for working with belief networks and IDs.
- **PrecisionTree** (<http://www.palisade.com>) performs decision analysis in Microsoft Excel using decision trees and IDs.
- **TreeAge Pro** (<http://www.treage.com>) enables the creation of decision trees, IDs and Markov models which can be analyzed with a sophisticated set of tools and offering a variety of graphs, charts and reports for communicating the results.

There are also some free tools with remarkable features. One of them is **GeNIe** (<http://genie.sis.pitt.edu>), with a very good user interface as well as efficient methods for computing and analyzing the strategies. Another one is **Elvira** (<http://leo.ugr.es/elvira>), which is mostly focused on research purposes although it offers a clear user interface and a complete set of algorithms for inference on IDs and Bayesian networks.

Therefore, it seems that IDs have been accepted as a model for dealing with decision problems in many different fields such as business, chemistry, environmental science, medicine, etc. There is a general agreement on their capability for offering a clear and powerful representation.

## 6 Summary and Discussion

IDs have proven to be a powerful tool for communicating ideas in decision-making problems. IDs represent a new dimension for decision models, Helfand and Pauker (1997). As an extension of Bayesian networks, IDs have benefitted from research in Artificial Intelligence, Decision Analysis and Statistics. The intersection of such disciplines has produced a useful instrument for representing and solving decision-making problems. The analysis presented here reveals complementary cross-discipline efforts and interactions.

We have studied the qualitative level of representation of an ID. Several new capacities have been identified. Unlike traditional (symmetric) IDs, intelligent modifications of IDs allow to directly model different kind of asymmetries –chance, decision and information asymmetries– in different ways. In Section 2, we have elaborated on conditional distribution trees, sequential decision diagrams, and sequential valuation networks, as the most outstanding approaches and mentioned a number of other graphical representations proposed in the literature. However, challenges to be able to model, e.g., constraints on the sequence of decisions remain.

Directly related to this and more generally, having different information constraints in all scenarios led us to study, in Section 3, the contributions made to the so-called sequencing constraint. Examples include diagnosis/troubleshooting problems. The main models here are the unconstrained IDs, the sequential valuation network and the sequential IDs, where the requirement of a directed path including all decision variables is dropped. Sequential influence diagram representation is very efficient. However, when the number of decisions without constraints is large, solving these models may be intractable. The limited memory influence diagrams besides drop the no-forgetting condition. The challenge here is perhaps to design (anytime) algorithms to find at any time approximate solutions when having time constraints that avoid finishing the whole exact algorithm.

We have discussed some literature and challenges associated with the use of continuous chance and decision variables in IDs. This is an area that is ripe for further research on the use of mixtures of truncated exponentials (MTEs) functions and mixtures of polynomials (MOPs) functions to approximate probability density functions. Using the extended Shenoy-Shafer architecture (Shenoy and West (2009a)), the MTE and MOP functions can be propagated in a strong junction tree to compute an optimal strategy.

We have not gone into other issues that are also important and are listed as follows:

- Decisions may involve *groups of people* rather than single decision makers, Poole (1997); Xiang (2002); Koller and Milch (2003); Detwarasiti and Shachter (2005); Ríos and Ríos-Insua (2006).
- Beyond the typical what-if questions on the quantitative values of the ID, *perturbations* of the ID may also include its *structure*, i.e. critical evaluation of functional relations on variables is also necessary. This also involves the complicated and burdensome process of generating an ID. Ideas here range from suggesting a case-based reasoning approach where candidate IDs are retrieved from a set of similar IDs (a case base), Lee et al. (2000), to using a neural network for yielding an initial ID that is later refined to generate a well-formed ID, Kim et al. (2002).
- *Explanation* capabilities of IDs are a key element that can improve the effectiveness and justification of decisions. Explanations enhance the decision maker's ability to understand and have confidence in the ID model, thereby accepting ID advices. Generating explanations is related to the search of the most relevant information and how to present it to the decision maker in a structured form. Although in Bayesian networks there has been much research in the topic, see a review in Lacave and Díez (2004), we expect to see more works in decision-making, Jimison et al. (1992); Bohanec and Rajkovic (1993); Gönül et al. (2006).
- Modeling *temporal* processes is one of the greatest challenges of IDs. *Dynamic IDs*, Tatman

and Shachter (1990), and *temporal IDs*, Provan and Clarke (1993), are possibilities that have in (partially observable) Markov decision processes a strong competitor.

- The need to deal with time- or modeling-pressured situations in developing *time-critical dynamic* decision support systems describes a realistic but difficult scenario. Xiang and Poh (1999, 2002) consider time-critical dynamic IDs to choose, from a meta-reasoning approach, the best model with optimal trade-off between model quality and computational tractability. Other authors construct anytime algorithms providing approximate strategies incrementally refined as computation progresses, Horsch and Poole (1996, 1998, 1999); Poh and Horvitz (1993); Ramoni (1995); Horvitz and Seiver (1997).
- A daunting challenge would be to face the continuous change in the environment. IDs are difficult to scale up and adapt to changing and evolving knowledge, policies, standards, regulations, operating procedures, etc. Perhaps IDs were not conceived to reach so far. However, combining the inclusion of time together with a mechanism that detects when the ID fails and suggests an alternative model is a possibility. These *adaptive* models have been the topic of many recent world conferences, like that organized by AAAI in 2005.
- Moreover, models need to effectively operate in the rapidly developing *distributed* infrastructures, gathering knowledge from different sources, Schneeweiss (2003), like from different computers in a network.

**Acknowledgments.** We are very grateful for discussions with T.D. Nielsen and J.Q. Smith in that beautiful isolated castle of the Scottish countryside (Gartmore) during the workshop on “Graphical Modelling of Dependent Uncertainties for Decision Support in Public Policy” in August 2004. Thanks also to T. Bedford who induced C. Bielza to prepare a talk in that workshop about challenges in large decision-making problems that finally ended up in the seed of this paper.

We are grateful to four DSS reviewers for constructive and useful comments on an earlier draft. The revised paper is much better as a result of their comments.

Research partially supported by the Spanish Ministry of Education and Science, projects TIN2007-62626 and TIN2007-67418-C03-03.

## References

Bielza, C., J.A. Fernández del Pozo, P. Lucas. 2008. Explaining clinical decisions by extracting regularity patterns. *Decision Support Systems* **44** 397–408.

- Bielza, C., M. Gómez, S. Ríos-Insua, J.A. Fernández del Pozo. 2000. Structural, elicitation and computational issues faced when solving complex decision problems with influence diagrams. *Computers and Operations Research* **27** 725–740.
- Bielza, C., P. Müller, D. Ríos. 1999. Decision analysis by augmented probability simulation. *Management Science* **45** 995–1007.
- Bielza, C., P.P. Shenoy. 1999. A comparison of graphical techniques for asymmetric decision problems. *Management Science* **45** 1552–1569.
- Bohanec, M., V. Rajkovic. 1993. Knowledge-based explanation in multiattribute decision making. *Computer aided decision analysis: Theory and applications*. Westport: Quorum Books, 189–204.
- Boutilier, C. 2005. The influence of influence diagrams on artificial intelligence. *Decision Analysis* **2** 229–231.
- Buede, D.M. 2005. Influence diagrams: A practitioner’s perspective. *Decision Analysis* **2** 235–237.
- Call, H.J., W.A. Miller. 1990. A comparison of approaches and implementations for automating decision analysis. *Reliability Engineering and System Safety* **30** 115–162.
- Charnes, J., P.P. Shenoy. 2004. Multi-stage Monte Carlo method for solving influence diagrams using local computation. *Management Science* **50** 405–418.
- Cinicioglu, E.N., P.P. Shenoy. 2009. Arc reversals in hybrid Bayesian networks with deterministic variables. *International Journal of Approximate Reasoning* **50** 763–777.
- Clemen, R.T. 1997. *Making Hard Decisions: An Introduction to Decision Analysis*. South-Western College Pub.
- Cobb, B.R., P.P. Shenoy. 2005a. Hybrid Bayesian networks with linear deterministic variables. F. Bacchus, T. Jaakkola, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 21st Conference*. AUAI Press, Corvallis, OR, 136–144.
- Cobb, B.R., P.P. Shenoy. 2005b. Nonlinear deterministic relationships in Bayesian networks. L. Godo, ed., *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 8th European Conference, ECSQARU 2005*. Lecture Notes in Artificial Intelligence 3571, Springer, Berlin, 27–38.
- Cobb, B.R., P.P. Shenoy. 2008. Decision making with hybrid influence diagrams using mixtures of truncated exponentials. *European Journal of Operational Research* **186** 261–275.

- Cobb, B.R., P.P. Shenoy, R. Rumi. 2006. Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. *Statistics & Computing* **16** 293–308.
- Cooper, G.F. 1988. A method for using belief networks as influence diagrams. *Proceedings of the 4th Conference on Uncertainty in Artificial Intelligence*. University of Minnesota, Minneapolis, 55–63.
- Covaliu, Z., R.M. Oliver. 1995. Representation and solution of decision problems using sequential decision diagrams. *Management Science* **41** 1860–1881.
- Demirer, R., P.P. Shenoy. 2006. Sequential valuation networks for asymmetric decision problems. *European Journal of Operational Research* **169** 286–309.
- Detwarasiti, A., R.D. Shachter. 2005. Influence diagrams for team decision analysis. *Decision Analysis* **2** 207–228.
- Dirac, P.A.M. 1927. The physical interpretation of the quantum dynamics. *Proceedings of the Royal Society of London, Series A* **113** 621–641.
- Fernández del Pozo, J.A., C. Bielza. 2002. An interactive framework for open queries in decision support systems. F.J. Garijo, J.C. Riquelme, M. Toro, eds., *Advances in Artificial Intelligence-IBERAMIA 2002*. Lecture Notes in Artificial Intelligence 2527, Springer, Berlin, 254–264.
- Fernández del Pozo, J.A., C. Bielza, M. Gómez. 2005. A list-based compact representation for large decision tables management. *European Journal of Operational Research* **160** 638–662.
- Fung, R.M., R.D. Shachter. 1990. Contingent influence diagrams. Technical report, Department of Engineering-Economic Systems, Stanford University, Stanford, Calif.
- Gómez, M. 2004. Real-world applications of influence diagrams. J.A. Gámez, S. Moral, A. Salmeron, eds., *Advances in Bayesian Networks*. Studies in Fuzziness and Soft Computing 146, Springer-Verlag, Berlin, 161–180.
- Gómez, M., C. Bielza, J.A. Fernández del Pozo, S. Ríos-Insua. 2007. A graphical decision-theoretic model for neonatal jaundice. *Medical Decision Making* **27** 250–265.
- Gönül, M.S., D. Önköl, M. Lawrence. 2006. The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems* **42** 1481–1493.
- Helfand, M., S.G. Pauker. 1997. Influence diagrams: A new dimension for decision models. *Medical Decision Making* **17** 351–352.

- Horsch, M.C., D. Poole. 1996. Flexible policy construction by information refinement. E. Horvitz, F. Jensen, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 12th Conference*. Morgan Kaufmann, San Francisco, CA, 315–324.
- Horsch, M.C., D. Poole. 1998. An anytime algorithm for decision making under uncertainty. G.F. Cooper, S. Moral, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 14th Conference*. Morgan Kaufmann, San Francisco, 246–255.
- Horsch, M.C., D. Poole. 1999. Estimating the value of computation in flexible information refinement. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 297–304.
- Horvitz, E., A. Seiver. 1997. Time-critical action: Representations and application. D. Geiger, P.P. Shenoy, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 13th Conference*. Morgan Kaufmann, San Francisco, CA, 250–257.
- Howard, R.A., J.E. Matheson. 1984. Influence diagrams. R.A. Howard, J.E. Matheson, eds., *Readings on the Principles and Applications of Decision Analysis, vol II*. Strategic Decisions Group, 719–762.
- Howard, R.A., J.E. Matheson. 2005a. Influence diagram retrospective. *Decision Analysis* **2** 144–147.
- Howard, R.A., J.E. Matheson. 2005b. Influence diagrams. *Decision Analysis* **2** 127–143.
- Howard, R.A., J.E. Matheson, M.W. Merkhofer, A.C. Miller, D.W. North. 2006. Comment on influence diagram retrospective. *Decision Analysis* **3** 117–119.
- Jensen, F., F.V. Jensen, D. Dittmer. 1994. From influence diagrams to junction trees. R.L. de Mantaras, D. Poole, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 10th Conference*. Morgan Kaufmann, San Francisco, CA, 367–373.
- Jensen, F.V., T.D. Nielsen, P.P. Shenoy. 2006. Sequential influence diagrams: A unified asymmetry framework. *International Journal of Approximate Reasoning* **42** 101–118.
- Jensen, F.V., M. Vomlelova. 2002. Unconstrained influence diagrams. A. Darwiche, N. Friedman, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 18th Conference*. Morgan Kaufmann, San Francisco, CA, 234–241.
- Jimison, H.B., L.M. Fagan, R.D. Shachter, E.H. Shortliffe. 1992. Patient-specific explanation in models of chronic disease. *Artificial Intelligence in Medicine* **4** 191–205.
- Kenley, C.R. 1986. Influence diagram models with continuous variables. Ph.D. thesis, Stanford University, Stanford, CA.

- Kim, J.K., K.C. Lee, J.K. Lee. 2002. Hybrid of neural network and decision knowledge approach to generating influence diagrams. *Expert Systems with Applications* **23** 237–244.
- Koller, D., B. Milch. 2003. Multi-agent influence diagrams for representing and solving games. *Games and Economic Behavior* **45** 181–221.
- Lacave, C., F.J. Díez. 2004. A review of explanation methods for heuristic expert systems. *Knowledge Engineering Review* **19** 133–146.
- Lauritzen, S.L., F. Jensen. 2001. Stable local computation with conditional Gaussian distributions. *Statistics and Computing* **11** 191–203.
- Lauritzen, S.L., D. Nilsson. 2001. Representing and solving decision problems with limited information. *Management Science* **47** 1235–1251.
- Lee, J.K., J.K. Kim, S.H. Kim. 2000. A methodology for modeling influence diagrams: A case-based reasoning approach. *International Journal of Intelligent Systems in Accounting, Finance & Management* **9** 55–63.
- Lerner, U., E. Segal, D. Koller. 2001. Exact inference in networks with discrete children of continuous parents. J. Breese, D. Koller, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 17th Conference*. Morgan Kaufmann, San Francisco, CA, 319–328.
- Liu, L., P.P. Shenoy. 2004. Representing asymmetric decision problems using coarse valuations. *Decision Support Systems* **37** 119–135.
- Madsen, A.L., F.V. Jensen. 1999. Lazy evaluation of symmetric Bayesian decision problems. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 382–390.
- Moral, S., R. Rumí, A. Salmerón. 2001. Mixtures of truncated exponentials in hybrid Bayesian networks. S. Benferhat, P. Besnard, eds., *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 6th European Conference, ECSQARU-2001*. Lecture Notes in Artificial Intelligence 2143, Springer, Berlin, 156–167.
- Murphy, K. 1999. A variational approximation for Bayesian networks with discrete and continuous latent variables. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 457–466.
- Nease, R.F., D.K. Owens. 1997. Use of influence diagrams to structure medical decisions. *Medical Decision Making* **17** 263–275.

- Nielsen, T.D., F.V. Jensen. 1999. Welldefined decision scenarios. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 502–511.
- Nielsen, T.D., F.V. Jensen. 2003. Representing and solving asymmetric decision problems. *International Journal of Information Technology and Decision Making* **2** 217–263.
- North, D.W., A.C. Miller, T. Braunstein. 1974. Decision analysis of intelligence resource allocation. Tech. Rep. DAH C15-73-C-0430, Stanford Research Institute, Menlo Park, CA.
- Olmsted, S.M. 1983. On representing and solving decision problems. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford, CA.
- Owens, D.K., R.D. Shachter, R.F. Nease. 1997. Representation and analysis of medical decision problems with influence diagrams. *Medical Decision Making* **17** 241–262.
- Pauker, S.G., J.B. Wong. 2005. The influence of influence diagrams in medicine. *Decision Analysis* **2** 238–244.
- Pearl, J. 2005. Influence diagrams-Historical and personal perspectives. *Decision Analysis* **2** 232–234.
- Poh, K.L., E.J. Horvitz. 1993. Reasoning about the value of decision-model refinement: Methods and application. D. Heckerman, A. Mamdani, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 11th Conference*. Morgan Kaufmann, San Francisco, CA, 174–182.
- Poland III, W.B. 1994. Decision analysis with continuous and discrete variables: A mixture distribution approach. Ph.D. thesis, Stanford University, Stanford, CA.
- Poland III, W.B., R.D. Shachter. 1993. Mixtures of Gaussians and minimum relative entropy techniques for modeling continuous uncertainties. D. Heckerman, A. Mamdani, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 9th Conference*. Morgan Kaufmann, San Francisco, CA, 183–190.
- Poole, D. 1997. The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence* **94** 7–56.
- Provan, G.M., J.R. Clarke. 1993. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15** 299–307.
- Qi, R., N.L. Zhang, D. Poole. 1994. Solving asymmetric decision problems with influence diagrams. R.L. de Mantaras, D. Poole, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 10th Conference*. Morgan Kaufmann, San Francisco, CA, 491–497.

- Ramoni, M. 1995. Anytime influence diagrams. *Proceedings of the IJCAI Workshop on Anytime Algorithms and Deliberation Scheduling*. American Association for Artificial Intelligence, 55–62.
- Rao, C.R. 1973. *Linear Statistical Inference and Its Applications*. 2nd ed. Wiley Series in Probability and Mathematical Statistics, Wiley.
- Ríos, J., D. Ríos-Insua. 2006. Negotiation over influence diagrams. Technical report, Rey Juan Carlos University.
- Schneeweiss, C. 2003. *Distributed Decision Making*. 2nd ed. Springer, New York.
- Shachter, R.D. 1986. Evaluating influence diagrams. *Operations Research* **34** 871–882.
- Shachter, R.D. 1999. Efficient value of information computation. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 594–602.
- Shachter, R.D., C.R. Kenley. 1989. Gaussian influence diagrams. *Management Science* **35** 527–550.
- Shachter, R.D., P.P. Ndilikilikisha. 1993. Using potential influence diagrams for probabilistic inference and decision making. D. Heckerman, A. Mamdani, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 383–390.
- Shenoy, P.P. 1992. Valuation-based systems for Bayesian decision analysis. *Operations Research* **40** 463–484.
- Shenoy, P.P. 2000. Valuation network representation and solution of asymmetric decision problems. *European Journal of Operational Research* **121** 579–608.
- Shenoy, P.P. 2006. Inference in hybrid Bayesian networks using mixtures of Gaussians. R. Dechter, T. Richardson, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 22nd Conference*. AUAI Press, Corvallis, OR, 428–436.
- Shenoy, P.P., J.C. West. 2009a. Inference in hybrid Bayesian networks with deterministic variables. C. Sossai, G. Chemello, eds., *Symbolic and Quantitative Approaches to Reasoning with Uncertainty-10th ECSQARU*. Lecture Notes in Artificial Intelligence 5590, Springer-Verlag, Berlin, 46–58.
- Shenoy, P.P., J.C. West. 2009b. Inference in hybrid Bayesian networks using mixtures of polynomials. Working Paper 321, University of Kansas School of Business, Lawrence, KS.
- Smith, J.E., S. Holtzman, J.E. Matheson. 1993. Structuring conditional relationships in influence diagrams. *Operations Research* **41** 280–297.

- Tatman, J.A., R.D. Shachter. 1990. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics* **20** 365–379.
- Xiang, Y. 2002. *Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach*. Cambridge University Press.
- Xiang, Y., K.L. Poh. 1999. Time-critical dynamic decision making. K.B. Laskey, H. Prade, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*. Morgan Kaufmann, San Francisco, CA, 688–695.
- Xiang, Y., K.L. Poh. 2002. Time-critical dynamic decision modeling in medicine. *Computers in Biology and Medicine* **32** 85–97.
- Zhang, N.L. 1998. Probabilistic inference in influence diagrams. G.F. Cooper, S. Moral, eds., *Uncertainty in Artificial Intelligence: Proceedings of the 14th Conference*. Morgan Kaufmann, San Francisco, CA, 514–522.