Causal Modeling Under a Belief Function Framework

THÈSE

(en cotutelle)

pour l'obtention du

Doctorat
de l'Université d'Artois & de l'Université de Tunis
Spécialité Informatique

par

Imen Boukhris

Composition du jury

Directeurs de thèse :	Salem Benferhat, Professeur, Université d'Artois, France Zied Elouedi, Professeur, ISG Tunis - Université de Tunis, Tunisie
Rapporteurs :	Philippe Leray, Professeur, Université de Nantes, France Aicha Mokhtari, Professeur, USTHB, Algérie
Examinateurs :	Lakhdar Sais, Professeur, Université d'Artois, France Eric Lefèvre, Maître de Conférences - HDR, Université d'Artois, France Nahla Ben Amor, Professeur, ISG Tunis - Université de Tunis, Tunisie

Centre de Recherche en Informatique de Lens – CNRS UMR 8188
Université d'Artois, rue Jean Souvraz, S.P. 18 F-62307, Lens Cedex France
Secrétariat : Tél.: +33 (0)3 21 79 17 23 – Fax : +33 (0)3 21 79 17 70
http://www.cril.fr

LAboratoire de Recherche Opérationnelle, de Décision et de Contrôle de processus
Institut Supérieur de Gestion, 41, avenue de la liberté, Le Bardo 2000, Tunis, Tunisie
Secrétariat : Tél.: +216 71 56 03 13 – Fax : +216 71 56 87 67
http://www.larodec.com

Abstract

This thesis focuses on the modeling of causality under the belief function framework. We have first analyzed and revised the theoretical foundations of existing associational belief models. Then, we have proposed a graphical structure that serves as a basis for the causal belief network. In this latter, uncertainty at the nodes is given in terms of conditional mass distributions. Since intervention is a crucial concept for an efficient causal analysis, we have introduced a counterpart of the do operator as a tool to represent interventions on causal belief networks. The proposed model allows computing the simultaneous effect of observations and interventions. It is assumed that considering an intervention as an action that always succeeds to force its target variable to have a precise value, by making it completely independent of its original causes, is a condition rarely achieved in real-world applications. Therefore, we have examined the treatment of interventions whose occurrence is imperfect and/or have imperfect consequences. In the last part of the thesis, we have proposed a model for causality ascription to interpret influential relationships between different attributes of the system namely causality, facilitation or justification in the presence observational and interventional data. Since decision makers are not only interested in ascribing causes, this model allows to define different strengths of a cause.

Résumé

La présente thèse s'intéresse à modéliser la causalité dans le cadre de la théorie des fonctions de croyance. Dans un premier temps, nous avons analysé et révisé les fondements théoriques des différents modèles associationnels crédibilistes existants. Nous avons proposé une structure graphique qui sert de base pour le réseau causal crédibiliste. Dans ce dernier, l'incertitude au niveau des nœuds est donnée en termes de distributions de masses conditionnelles. Comme la notion d'intervention est d'une importance capitale pour une analyse causale efficiente, nous avons introduit la contrepartie de l'opérateur "do" pour la représentation des interventions dans les réseaux causaux crédibilistes. Le modèle proposé permet ainsi de calculer l'effet simultané des observations et des interventions. Il est admis que considérer l'intervention comme une action qui réussit toujours à mettre sa cible à une valeur précise en la rendant complètement indépendante de ses causes originales est une condition rarement réalisée dans les applications réelles. De ce fait, nous avons examiné le traitement des interventions dont l'occurrence et/ou les conséquences sont imparfaites. Dans la dernière partie de la thèse, nous avons proposé un modèle d'attribution de causalité permettant d'interpreter les relations d'influence qui existent entre les différents attributs du système à savoir la causalité, la facilitation ou encore la justification en présence de données observationnelles et interventionnelles. Motivé par le fait que les preneurs de décision ne sont pas uniquement intéressés par l'attribution de la causalité, ce modèle permet de définir différentes forces d'une cause.

Acknowledgements

Carrying out a Ph.D. is transforming a laborious project into an enjoyable experience. It is just like climbing a high peak, step by step, accompanied with disappointment, hardships, frustration, encouragement and trust and with so many people's kind help. When I found myself at the top savoring the beautiful scenery, I realized that it was teamwork that got me there. I would like to express my gratitude to all those people who helped me.

My sincere thanks to my supervisors, Pr. Zied Elouedi (University of Tunis, Tunisia) and Pr. Salem Benferhat (University of Artois, France), who without any hesitation, accepted me as their Ph.D student. From that time forward, they offered me so much advice and patiently supervised me. I have learned a lot from them, without their help I could not have finished my dissertation successfully. It is not sufficient to express my gratitude with only a few words.

I am also thankful to all members of my thesis committee: Pr. Philippe Leray (University of Nantes, France), Pr. Aicha Mokhtari (University of Sciences and Technology Houari Boumedine, Algeria), Pr. Lakhdar Sais (University of Artois, France), Dr. Eric Lefèvre (University of Artois, France) and Pr. Nahla Ben Amor (University of Tunis, Tunisia).

My thanks go to all members of my laboratory *LARODEC* and all members of my laboratory *CRIL*.

Finally, I convey my heartfelt thanks to all the persons who have supported me and have offered me an ideal environment in which I could concentrate on my research.

Contents

List of Figures

List of Tables

General Introduction

Causality plays an important role in many fields, from physics to medicine to artificial intelligence. Indeed, causal knowledge simplifies decision-making. In fact, it enables to anticipate the dynamics of events when the system is evolving and thus choose the right actions to achieve the goals.

Though at first glance, it seems obvious what causation is, there are difficulties to define which event truly causes another. This makes difficult the comprehension and modeling of causality. Actually, it can be seen as a regular association (Hume, 2006), a counterfactual dependence (Lewis, 2004) or a probability raising (Eells, 1991).

A general agreement says that causal links should be well distinguished from statistical relations. Indeed, correlations are symmetric relationships where the occurrence of two events is observed at the same time but an external action on one of them by forcing it to take place will not affect the other event. However, a paradigmatic assertion in causal relations is that the manipulation of a genuine cause will result in the variation of an effect.

Motivation

Researchers in the Artificial Intelligence (AI) field were interested in the problems arising from the modeling of causality (e.g. (Shafer, 1996; Pearl, 2000; Halpern & Pearl, 2005)) motivated by the fact that it is important to provide the systems of inference or decision-making with explanation capacity for an operator or human user.

Graphical models are efficient and simple ways for representing and reasoning under uncertainty since they compactly represent dependance relations. Indeed, Bayesian networks (Pearl, 1988) are popular within the AI community. In these directed acyclic graphs, edges represent conditional relationships. However, the direction of the edges is induced by the order in which variables are considered without necessarily reflecting causal links. Besides, since they are based on observational data, several networks are equivalent according to the Markov property (i.e., they encode the same joint distribution). Only one of them models cause/effect relationships: the causal Bayesian network.

Hence, causal Bayesian networks are an extension of classical Bayesian networks where

links between variables follow the causal process allowing a more informative analysis. In fact, through these networks it is possible to model causal knowledge which is usually uncertain. It is also possible to predict not only the spontaneous behavior of the system but also the effect of interventions. While conditioning is used to compute the effect of observations, the do operator (Pearl, 2000) is used to compute the impact of external action.

Nevertheless, probabilistic graphical models are effective when a very complete statistical knowledge description of the modeled system is available. In fact, a probability distribution, as good as it is, does not distinguish between equiprobability and ignorance situations. Besides, probability theory cannot handle situations where experts are unable to provide the whole numerical values. Thus, a reasoning based on Bayesian networks may be misleading is some situations. In fact, possibilistic networks (Fonck, 1994; Ben Amor et al., 2003) are adequate in cases requiring pure qualitative and ordinal handling and belief networks (Ben Yaghlane et al., 2003; Xu & Smets, 1994) are ideal tools in situations of ignorance and incomplete knowledge.

In this thesis, we choose to work with the belief function theory since it represents an appropriate framework for experts to express their beliefs in a flexible way. Indeed, the theory of belief functions is now well established as a general framework for reasoning with uncertainty, and has well understood connections to other frameworks such as probability, possibility and imprecise probability theories.

Belief function networks with conditional dependencies were first introduced with Shenoy (1993) in valuation based systems (VBS), called valuation networks. Xu et al. (1994) have presented an alternative framework to the general VBS, called evidential network with conditional belief functions (ENC). In this network, relations between the variables in the network are defined per edge. To overcome this limitation, Ben Yaghlane et al. (2003) have proposed the Directed EVidential Network with conditional belief function (DEVN) where they allow to define conditionals either for one parent node of for all parents. These networks have the same structure as a Bayesien network. However, the manner in which conditional beliefs are defined is different from that one in which conditional probabilities are traditionally defined in Bayesien networks: each edge in the graph represents a conditional relation between the two nodes it connects. Recently, Simon et al. (2008) proposed another model called Evidential Network (EN) where conditional dependencies are defined given all the parents. There is no distinction between this network and a Bayesian network in terms of representation and propagation of beliefs.

The aim of this thesis is to model causality under the belief function framework. In fact, despite the representation power of this theory and the importance of causality, no works have been presented in this area. Causal networks provide information about the dynamics of the system under study. However, the definition of a causal network is based on the definition of its corresponding associational networks according to the uncertainty theory. Thus, before modeling causality, we have to analyze the panoply of existing associational belief networks to know which one could be appropriate to serve as a basis to our causal network under the belief function theory.

Contributions

This thesis contains four important and new contributions regarding : foundations of associational belief networks, definition of causal belief networks, modeling different forms of non-standard interventions that may have imperfect consequences and causality ascription.

1. Associational belief networks: we first analyze the theoretic foundations of belief networks. We show that contrary to expectations, even though the belief function theory is a generalization of probability theory, the belief network fails to collapse into a Bayesian one. This is due to the fact that in these networks, conditional distributions can be defined per edge. Indeed, belief networks can be defined either per edge or for all parent nodes. Then, to add more flexibility we propose a new belief network where conditional distributions are defined for one or more parent nodes. To compute the global joint distribution, local distributions should be aggregated. Accordingly, we clarify the computation of the global joint distribution by explaining why, how and when extensions should be made. Finally, we propose new definitions of these extensions (for the vacuous and the ballooning concepts). These new rules departs from existing ones. In fact, instead of being based on a least committed transfer of masses, they are based on a uniform one. Doing such a way is more appropriate to reflect the initial knowledge.

2. Definition of causal belief networks: the belief function theory is adequate to formalize imperfect causal knowledge that agents usually possess especially cases of ignorance. Accordingly, we need a graphical structure to simply represent and reason from such causal knowledge. Based on our associational belief network, we introduce causal belief network where arcs are interpreted as causal links. On this network, we can compute the effects of observations and also those of external actions. For that we propose a counterpart the "do" operator under the belief function framework.

   We explain that after an intervention, the state of the target variable concerned by the manipulation is therefore totally dependent of this external action and independent of its original causes. Accordingly, we study two different equivalent approaches to deal with interventions: graph mutilation and graph augmentation. The first way is to interpret an intervention by cutting off the edges linking it to its parents. The rest of the network remains unchanged. The resulting graph is a mutilated causal belief network. Another alternative is to add a new fictive variable, the "DO" variable, as a parent node of the variable concerned by an intervention. The resulting graph is called an augmented causal belief network. Then, we show that even if the computation of the belief joint distribution is very different from the one used in probability theory the graph augmentation and graph mutilation are equivalent methods and lead to the same joint distribution.

3. Modeling different forms of non-standard interventions: we point out that considering the intervention as a *certain* action which always succeeds to put its target at a *precise* value by making it completely independent of its original causes is a condition that is rarely achieved in real world applications. Indeed, we show that an intervention can be uncertain or imprecise in the sense that it may imperfectly occur. Then, we explain that even if it takes place with a degree of belief, it can have imperfect consequences which means that it may not succeed to put its target into one specific value. Furthermore, an intervention can interact with the original causes of the target variable. All these

different types of interventions should be well modeled. In fact, the methods proposed for certain and precise interventions will lead to erroneous decisions if they were applied in these cases. To tackle this problem, we provide methods to deal with these non-standard interventions. We argue that our methods are very naturally encoded with causal belief networks.

4. Causality ascription: causal discovery is a challenging task and ascribing the causes of an abnormal event occurring in the normal course of things is a very useful task in many applications (e.g., intrusion detection problems, medical applications). We first present a belief causality ascription model that allows an agent to discriminate between potential causes of an abnormal event based on his imperfect background knowledge. This latter, can be represented by a causal belief network and a sequence of reported events. To achieve this goal, we provide definitions of the concepts of acceptance, rejection and ignorance. Then, we explain that by defining more levels of these notions, we will be able to ascribe the strength of the causal connection. Besides, we explain how facility and justification will be distinguished from causality.

   Unlike the qualitative method based on nonmonotonic consequence relations, our method is appropriate to handle n-ary variables. Moreover, we show that attenuation and confirmation totally make sense which is not the case for the existing qualitative model. In fact, if an event is held as accepted, then after the observation of a second event, it only may remain accepted or becomes rejected and cannot be attenuated or confirmed.

   Finally, to avoid detecting spurious correlations, we propose to use not only observational data but also interventional data.

Organization

Our thesis is organized in six chapters distributed into two parts. Note that illustrative examples are presented in the different chapters.

Part I: Theoretical aspects composed of two chapters as follows:

- Chapter 1 gives the necessary background regarding the basic concepts of belief function theory. A brief overview of some graphical models where the knowledge is formalized with belief functions is provided at the end of this chapter.

- Chapter 2 explains the basic ideas of causal knowledge that are used throughout our work. Then, a more attention is given to causality ascription methods and to existing causal networks. It also examines an important concept in causality namely the treatment of interventions.

Part II: Modeling causality under the belief function framework composed of four chapters as follows:

- Chapter 3 in its first part investigates the case of belief networks in which beliefs are Bayesian. The second part is dedicated to the analysis and clarification of existing belief function networks. Then, we define a graphical structure called belief network with conditional beliefs where uncertainty is given in terms of conditional a priori masses.

The computation of the joint distribution is based on new definitions of some belief function concepts useful to handle belief networks.

- Chapter 4 presents our causal model called causal belief network where uncertainty at nodes is represented with conditional a priori mass distributions. The proposed model allows to truly identify causal links through the notion of interventions. This network is based on the graphical structure defined in Chapter 3, where the arrows are interpreted as causal links. The treatment of interventions in both mutilated and augmented graphs is detailed and the equivalence between these methods is verified. Furthermore, we emphasize that interventions are imperfect in real-world applications. We explain how to compute their effects at the end of this chapter.

- Chapter 5 proposes a model that an intelligent agent will use to ascribe causality from a sequence of observations or interventions occurring in his environment under the belief function framework. To model such changes, we use the concepts of acceptance and rejection instead of changes in uncertainty distributions. We introduce definitions of acceptance and rejection allowing the categorization of causes according to their strength.

- Chapter 6 presents all necessary elements which are needed to describe the proposed models. Thus, we first develop methods to implement our proposed causal belief network and to simulate the effect of an intervention using the belief graph mutilation and augmentation approaches. Then, we show the usefulness of the belief causality ascription model to ascribe the causes of an abnormal event.

Finally, a general conclusion gives a summary of the results achieved in this thesis and presents possible future developments.

Two appendices complete this thesis. They provide proofs of propositions given respectively in Chapter 4 and Chapter 5.

Part I

Theoretical aspects

Chapter 1

Belief Function Theory

1.1 Introduction

Reasoning under uncertainty either quantitative or qualitative reasoning is required in most real world applications. Uncertain reasoning is dominated by probabilistic tools. This is somewhat surprising, since many situations involving imperfection (uncertainty, imprecision, incompleteness or vagueness) cannot be represented in an appropriate way within the classical probability framework. Accordingly, a modeler must be open to all the available tools, using the right one for the right problem. In this thesis we are interested in the belief function theory. It was developed to deal with imprecise and uncertain information. It is ideal in situations of ignorance and incomplete knowledge.

The belief function theory, known also as the theory of evidence (Barnett, 1981), (Guan & Bell, 1991) or Dempster-Shafer theory (Gordon & Shortliffe, 1984) is a general framework for reasoning with uncertainty. It shows its efficiency in many real-world applications (e.g., multi-sensor fusion (Kim & Swain, 1995; Appriou, 1999), pattern recognition (Tupin et al., 1999; Denœux & Zouhal, 2001), classification (Elouedi et al., 2001; Trabelsi et al., 2011), image processing (Bloch, 1996; Lefevre et al., 2000), system analysis (Simon et al., 2008), threat assessment (Benavoli et al., 2009) as well as environmental monitoring (Ben Abdallah et al., 2012)).

Originally introduced by Arthur Dempster (1967) as a special case of upper and lower probabilities, this theory is considered as a generalization of the Bayesian inference when there is not an a priori on the parameters (Dempster, 1968).

Glenn Shafer (1976) offered a re-interpretation of Dempster's work and coined this formalism the term "belief functions". He attached likelihood to events. He presented the belief function theory as a general framework to represent uncertainty.

The term belief function theory is used in the literature according to two family of models, namely:

- Models based on probability theory: The Dempster-Shafer model (Dempster, 1967), the lower probability model (Walley, 1991), the theory of hints

(Kohlas & Monney, 1995).

- Non-probabilistic model: The transferable belief model (TBM) (Smets, 1988a, 1993a, 1988b) represents quantified beliefs built without requiring any underlying hidden probability measures (Smets, 1995b). It extends the model proposed by Shafer. It is based on a subjective interpretation under which beliefs can be entertained outside any decision context. Accordingly, two levels are distinguished:

  – a credal level ("credo" meaning I believe), in which beliefs are quantified by belief functions (those introduced by Shafer (1976)) (static part). The reasoning process on these beliefs and their revision according to the disposal of a new piece of information is also made at this level (dynamic part).

  – a pignistic level ("pignus" means a bet) at which decisions are made.

  Relations between belief functions held at the credal level and probabilities held at the pignistic level are given in (Smets, 1990b).

In this thesis, we deal with the interpretation of the belief function theory as explained by the TBM. This chapter is organized as follows: we start, in Section 1.2, by describing different belief functions that are used to represent knowledge under the belief function framework. Then, in Section 1.3 several basic operations are detailed like combination, conditioning or discounting. Section 1.4 is dedicated to multi-variable operations and Section 1.5 explains how to decide using the pignistic transformation. In Section 1.6, we explain how graphical structures can simply and compactly describe a given system.

1.2 Belief function theory: representation

Let us denote by Θ the finite non empty set including n elementary events (hypotheses) representing the solutions of a given problem. These events are assumed to be exhaustive and mutually exclusive. The set Θ, called the frame of discernment, is defined as:

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$$

We handle events on the powerset of Θ, denoted by 2^Θ. This set contains singleton hypotheses of Θ, all possible disjunctions of these hypotheses as well as the empty set. In the following, we will denote by θ a singleton hypothesis and by A a proposition or an event designating either a hypothesis or a disjunction of hypotheses.

$$2^\Theta = \{A, A \subseteq \Theta\} = \{\emptyset, \{\theta_1\}, \ldots, \{\theta_n\}, \{\theta_1, \theta_2\}, \ldots, \Theta\}$$

Example 1.1. *A murder has been committed. There are three suspects* Θ = {*Mary, John, Peter*}. *The corresponding powerset of* Θ *is composed as follows:*

2^Θ = {$\emptyset$, {*Mary*}, {*John*}, {*Peter*}, {*Mary, John*}, {*Mary, Peter*}, {*Peter, John*}, {*Mary, Peter, John*}}.

1.2.1 Basic belief assignment

Within the belief function theory (Shafer, 1976), beliefs are expressed on propositions belonging to the powerset of Θ.

The basic belief assignment (*bba*), denoted by m^{Θ}, is a mapping from 2^{Θ} to [0,1] such that any proposition is associated with a real number belonging to [0,1] where the sum overall subsets is equal to 1. When there is no ambiguity, m^{Θ} will be shortened m.

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{1.1}$$

The value $m(A)$ is a basic belief mass (*bbm*) assigned to A. It is defined as (Cooke & Smets, 2001):

- the mass that is *exactly committed to the event* A of Θ, and due to a lack of information, cannot be allocated to any strict subset of A. Hence, it does not support any $B \subset A$,
- a mass that could freely be given to any subsets of A if we were given new information.

Note that a *bbm* associated with a proposition $A \subseteq \Theta$, $A = \cup_{B \in \Theta} B$ does not give any information about the belief on Bs (i.e., subsets composing A). In fact, unlike probability theory where:

$$P(A) = \sum_{\cup_{B \in \Theta}} P(B)$$

The additivity property is not satisfied (Klir & Wierman, 1998):

$$m(A) \neq \sum_{\cup_{B \in \Theta}} m(B) \tag{1.2}$$

The subsets of Θ such that m(A)$>$ 0 are called focal elements. The union of all focal elements is called its core. Shafer (1976) initially does not consider the empty set as a focal element (i.e., the impossible proposition). The mass function satisfying this constraint is called *normalized*:

$$m(\emptyset) = 0$$

In the TBM, Smets removes this constraint and allows unnormalized belief functions. In fact, he considers that the frame of discernment may be not exhaustive and accordingly m($\emptyset$) quantifies the agent's belief that the true answer is not in Θ carrying the idea that the chosen model might not fit reality with enough precision.

The normalization process is defined as follows:

$$m(A) = \begin{cases} 0 \text{ if } A = \emptyset \\ K \cdot m(A) \text{ otherwise} \end{cases} \tag{1.3}$$

where K^{-1} = 1 - m($\emptyset$). It is called the normalization factor.

Since the belief function theory models several types of imperfection, special *bba*s were defined. In particular, we have:

- Certain *bba*: the case where there is exactly one focal element that is a singleton.

  $m(\theta) = 1$ for one particular element of Θ

- Bayesian *bba*: the case where all focal elements are singletons, the particular case of probabilities.

  if $m(A) > 0$ then $|A| = 1$, where $|A|$ stands for the cardinal of A.

- Consonant *bba*: the case where all focal elements are nested. It is the special case of possibility theory (Dubois et al., 2001).

- Vacuous *bba*: the case where the normalized *bba* models the state of the total ignorance. Θ is the unique focal element, defined such that (Shafer, 1976):

  $m(\Theta) = 1$

- Categorical *bba*: the case where the *bba* has a unique focal element A:

  $m(A) = 1, A \subseteq \Theta$

- Simple support function (*ssf*) *bba*: the case where the evidence only supports a subset A of Θ, i.e., focal elements are $\{A, \Theta\}$.

- Non-dogmatic *bba*: the case where the frame of discernment (Θ) is a focal element:

  $m(\Theta) > 0$

Example 1.2. *(continued)*
Assume that $\Theta = \{Mary, John, Peter\}$. *Witnesses may express their beliefs about the murderer in a flexible way by supporting one or more than one proposition as shown in Table 1.1:*

Table 1.1: Example of *bba*s expressed on the frame of discernment Θ

	certain	categorical	Bayesian	consonant	vacuous	ssf	any
$\emptyset$	0	0	0	0	0	0	0.1
{Mary}	0	0	0.2	0.4	0	0	0.05
{John}	0	0	0.7	0	0	0	0.14
{Peter}	1	0	0.1	0	0	0	0.2
{Mary,John}	0	0	0	0.5	0	0	0.3
{Mary,Peter}	0	0	0	0	0	0.7	0.11
{John,Peter}	0	1	0	0	0	0	0.01
Θ	0	0	0	0.1	1	0.3	0.09

1.2.2 Belief function

The belief of an agent that the actual world lies in A can be represented by the belief function, *bel*: $2^\Theta \rightarrow [0,1]$, defined as the sum masses of the belief committed exactly to A (i.e., $m(A)$) and to every proper subsets of A. It represents the *total belief* that one commits to A without being also committed to $\overline{A}$.

It is considered as a *justified specific* support, since the *bbm* assigned to the empty set supporting A and at the same time $\overline{A}$ is not taken into consideration for the computation of $bel(A)$ neither for the computation of $bel(\overline{A})$.

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad \text{and} \quad bel(\emptyset) = 0 \tag{1.4}$$

The basic belief assignment can be recovered from the belief function as follows:

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} bel(B) \tag{1.5}$$

Another function, related to *bel* is *doubt*: $2^{\Theta} \rightarrow [0,1]$, it represents the extent to which one disbelieves a proposition (or believes its complement). It is defined as:

$$doubt(A) = bel(\overline{A}) \tag{1.6}$$

The belief function can be characterized without reference to the mass function. Shafer (1976) notes that a belief function should satisfy the following rules:

$$\begin{aligned} bel(\emptyset) &= 0 \\ bel(\Theta) &= 1 \\ bel(A_1 \cup \ldots \cup A_n) &\geq \sum_{\emptyset \neq I \subset \{1,\ldots,n\}} (-1)^{|I|+1} bel(\bigcap_{i \in I} A_i), \\ &\forall n > 0, \forall A_1 \cup \ldots \cup A_n \subseteq \Theta \end{aligned} \tag{1.7}$$

1.2.3 Plausibility function

The plausibility function *pl*, the dual measure of *bel*, represents the extend to which one fails to doubt the proposition A. It quantifies the *maximum amount of belief* that could be given to a subset A. It represents the *potential specific* support since the *bbm* included in $pl(A)$ could be transferred to non empty subsets of A if some new information could justify such a transfer.
pl: $2^{\Theta} \rightarrow [0,1]$ is defined as the sum of masses compatible with A (i.e., do not contradict A).

$$pl(A) = bel(\Theta) - bel(\overline{A}) = 1 - m(\emptyset) - doubt(A) \tag{1.8}$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{1.9}$$

The basic belief assignment can be recovered from the plausibility function as follows:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B+1|} pl(\overline{B}) \tag{1.10}$$

1.2.4 Commonality function

The value $m(A)$ is the part of belief exactly committed to $A \subseteq \Theta$. In the light of a new piece of information, it may be transferred to any subset of A. The commonality function q: $2^{\Theta} \rightarrow [0,1]$, measures the total mass that can move freely to every point of a proposition A (Shafer,

1976). It is computed as the sum of the masses allocated to the supersets of A (i.e., having A in common).

$$q(A) = \sum_{B \supseteq A} m(B) \tag{1.11}$$

The basic belief assignment can be recovered from the commonality function as follows:

$$m(A) = \sum_{A \subseteq B} (-1)^{|B-A|} q(B) \tag{1.12}$$

1.2.5 Properties of belief functions

The plausibility function is related to the belief function as shown in Figure 1.1. It can be expressed as:

$$pl(A) = bel(A) + \sum_{A \cap B \neq \emptyset, B \not\subset A} m(B) \tag{1.13}$$

Accordingly,

$$pl(A) \geq bel(A) \tag{1.14}$$

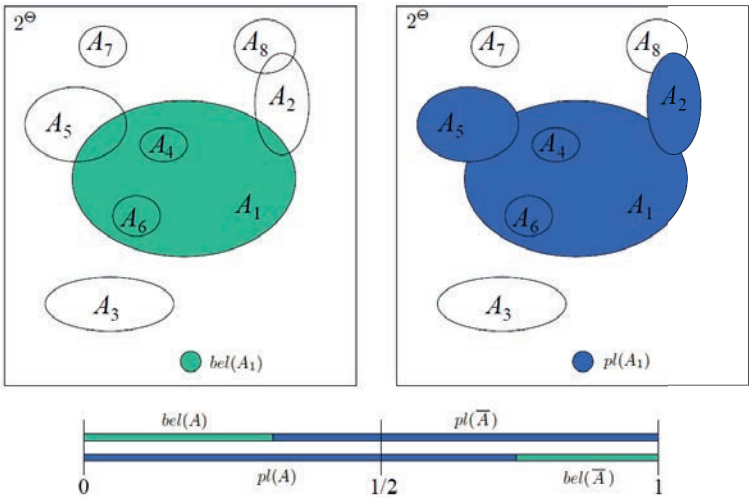


Figure 1.1: Belief function measures

Non-additivity

Unlike probability theory where increasing the beliefs on a proposition A, implies the decrease of the beliefs on $\overline{A}$ (i.e., $P(A) + P(\overline{A}) = 1$), under the belief function framework, a belief on a proposition does not give any information about the belief on the complement.

Belief function sub-additivity

$$bel(A) + bel(\overline{A}) \leq 1 \tag{1.15}$$

Plausibility function over-additivity

$$pl(A) + pl(\overline{A}) \geq 1 \tag{1.16}$$

Monotonicity

If two events A and B are such that $A \subseteq B$, then:

$$\begin{array}{l} bel(A) \leq bel(B) \\ pl(A) \leq pl(B) \end{array} \tag{1.17}$$

This relation reflects the fact that if A occurs less often than B (because B contemplates more occurrences), then the belief (resp. plausibility) of A must be less than the belief (resp. plausibility) of B.

1.3 Belief function theory: basic operations

Reasoning about knowledge can be seen as an operation involving the mass transfer from one subset of Θ to another.

1.3.1 Conjunctive rule of combination

The conjunctive rule of combination is used to combine two reliable and distinct sources of information. The sources are assumed to provide pieces of evidence m_1 and m_2 that may be combined to give one resulting mass $m_1 \textcircled{\cap} m_2$. It is defined as the orthogonal sum of two *bba*s m_1 and m_2, whose focal elements are all the possible intersections between pairs of focal elements of m_1 and m_2 respectively.

$$m_1 \textcircled{\cap} m_2(A) = \begin{cases} \sum\limits_{B \cap C = A} m_1(B) \cdot m_2(C) & \text{if } A \neq \emptyset,\ \forall B, C \subseteq \Theta \\ 0 & \text{otherwise} \end{cases} \tag{1.18}$$

The mass allocated to the empty set may be seen as the degree of conflict between the two sources.

Note that the use of commonality functions simplifies the computation of this rule since it is reduced to a simple product as follows:

$$q_1 \textcircled{\cap} q_2(A) = q_1(A) \cdot q_2(A). \tag{1.19}$$

This rule of combination has the following properties:

- Associative: $(m_1 \textcircled{\cap} m_2) \textcircled{\cap} m_3 = m_1 \textcircled{\cap} (m_2 \textcircled{\cap} m_3)$
- Commutative: $m_1 \textcircled{\cap} m_2 = m_2 \textcircled{\cap} m_1$
- Non-idempotent: $m_1 \textcircled{\cap} m_1 \neq m_1$
- Having the vacuous *bba* as a neutral element: $m_1 \textcircled{\cap} m_0 = m_1$, m_0 denotes here a vacuous *bba*.

If the normalization is performed, then the conjunctive rule of combination is reduced to the *Dempster rule of combination* denoted by $\oplus$ and defined as:

$$m_1 \oplus m_2(A) = \begin{cases} \dfrac{m_1 \textcircled{\cap} m_2(A)}{1 - m_1 \textcircled{\cap} m_2(\emptyset)} & \text{if } A \neq \emptyset,\ \forall A \subseteq \Theta \\ 0 & \text{otherwise} \end{cases} \tag{1.20}$$

Thanks to the associative and commutative properties, several sources of information are combined by applying repeatedly the chosen rule as shown in Figure 1.2.

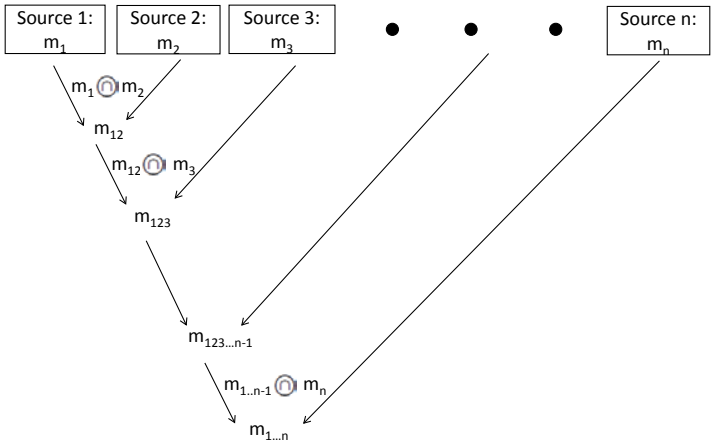


Figure 1.2: Combining n sources

1.3.2 Disjunctive rule of combination

The disjunctive rule of combination, denoted $\textcircled{\cup}$, is used to combine two *bba*s m_1 and m_2 when at least one piece of evidence holds. It is defined as:

$$m_1 \textcircled{\cup} m_2(A) = \sum_{B \cup C = A} m_1(B) \cdot m_2(C), \forall B, C \subseteq \Theta \tag{1.21}$$

As the conjunctive rule of combination, this rule is commutative, associative and non-idempotent. Its neutral element is the *bba* which assigns the total mass of belief to the empty set, i.e., $m(\emptyset) = 1$.

1.3.3 Decombination

The decombination operator (Smets, 1995a) (also called the removal operator (Shenoy, 1994)) is denoted by $\textcircled{\emptyset}$ (and $\ominus$ for its normalized form). It allows to remove a non-dogmatic *bba* m_2 from a combined mass $m_1 \textcircled{\cap} m_2$ in order to find m_1. Formally,

$$(m_1 \textcircled{\cap} m_2) \textcircled{\emptyset} m_2 = m_1$$

The commonality function is an interesting and useful tool, simplifying the computations for the decombination operation by reducing it to a pointwise division. The decombination

operator can therefore be simply computed with the commonality function as follows:

$$q_1 \textcircled{\not} q_2(A) = \frac{q_1(A)}{q_2(A)} \tag{1.22}$$

1.3.4 Dempster's rule of conditioning

Smets (1990a) qualified Dempster's rule of conditioning as one of the natural ingredients and the center of the transferable belief model.

Let us suppose that an agent allocates a mass to a proposition A. Conditioning allows to change the knowledge the agent had after the information that an event $B \subseteq \Theta$ is true. All non vacuous propositions implying $\overline{B}$ will be transferred to the part of A compatible with the evidence namely, $A \cap B$.

In the case, where $A \cap B = \emptyset$, several methods exist for transferring the remaining evidence (Smets, n.d.). $m(A|B)$ denotes the degree of belief of A in the context where B holds with $A, B \subseteq \Theta$. The Dempster rule of conditioning is computed by:

$$m(A|B) = \begin{cases} K \cdot \sum\limits_{C \subseteq \overline{B}} m(A \cup C) & \text{if } \ A \subseteq B, A \neq \emptyset \\ 0 \text{ if } \ A \nsubseteq B \end{cases} \tag{1.23}$$

where $K^{-1} = 1 - m(\emptyset \mid B)$.

Since m, bel, pl are in one-to-one correspondence, we get:

$$bel(A|B) = \frac{bel(A \cup \overline{B}) - bel(\overline{B})}{1 - bel(\overline{B})} \tag{1.24}$$

$$pl(A|B) = \frac{pl(A \cap B)}{pl(B)} \tag{1.25}$$

1.3.5 Relationship between Dempster's rule of combination and Dempster's rule of conditioning

The conjunctive rule of combination rule has a simple expression using the unnormalized Dempster's rule of conditioning (Dubois & Prade, 1986), (Smets, 1993c):

$$f_1 \oplus f_2(A) = \sum_{B \subseteq \Theta} f_1(A|B) \cdot m_2(A), \quad \text{where} \quad f \in [m, bel, pl, q] \tag{1.26}$$

Thus, when $f = m$ we have:

$$m_1 \oplus m_2(A) = \sum_{B \subseteq \Theta} m_1(A|B) \cdot m_2(A) \tag{1.27}$$

The conditioning rule can be defined as a special case of the rule of combination where one source is certain (note that it may be imprecise). However, the underlying intuitions are completely different. One is concerned with belief updating, while the other is concerned with belief combination. In fact, through conditioning the masses originally assigned to A will be

updated following the disposal of a new more precise information saying that the proposition B is certain (more precisely a categorical *bba* focused on B), i.e., $m(B) = 1$.

$$m(A|B) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A} m(C)}{pl(B)} \text{ if } \ A \subseteq B, A \neq \emptyset \\ 0 \ \text{ otherwise} \end{cases} \tag{1.28}$$

1.3.6 Discounting

A basic belief assignment can be weakened before the combination to take into account the reliability of an expert by the discounting method defined as:

$$m^{\alpha}(A) = \begin{cases} (1-\alpha) \cdot m(A), \forall A \subset \Theta \\ \alpha + (1-\alpha) \cdot m(A), \ \text{ if } A = \Theta \end{cases} \tag{1.29}$$

The discounting operation is controlled by a *discount rate* α taking values between 0 and 1. If $\alpha = 0$, the source is fully reliable and the belief function is unchanged; if $\alpha = 1$, the belief function is transformed into the vacuous belief function, meaning that the information provided by the expert is completely discarded.

1.4 Belief function theory: multi-variable operations

We have presented the basic concepts of the belief function theory. These mechanisms are based on the assumption that *bba*s are defined on the same frame of discernment. However, this constraint limits the practical applications. Let us consider in what follows, a first frame Θ and a second frame Ω.

We will present multi-variable operations on these frames, namely the vacuous extension and its inverse operation: marginalization, as well as the deconditioning operation called the ballooning extension. These operations are based on the principle of minimal commitment.

1.4.1 Principle of minimal commitment

Like the maximum entropy principle and the principle of minimum specificity (Dubois & Prade, 198), this principle (Hsia, 1991; Smets, 1993b) reflects a form of conservatism in the allocation of beliefs. It indicates that, given two belief functions compatible with a set of constraints, the most appropriate supported belief is the least committed one (i.e., the least informative one). It formalizes the idea: *one should never give more support than justified to any subset of* Θ.

The commitment of a belief function can be defined according to two approaches:

- A quantitative approach using some measures of uncertainty, e.g., non-specificity measure (Dubois & Prade, 198) defined as:

$$N(A) = \sum_{\emptyset \neq A \subseteq \Theta} m(A) \cdot log_2(A) \tag{1.30}$$

- An ordinal approach called also the principle of maximal plausibility (Dubois & Prade, 198; Smets, 2000). A plausibility function pl_2 is less committed than pl_1, if there is a strict inequality such that:

$$pl_1 < pl_2 \tag{1.31}$$

Accordingly, the least committed belief function is the vacuous belief function, i.e., $\mathrm{m}(\Theta) = 1$.

1.4.2 Cylindrical extension and projection

Cylindrical extension

The cylindrical extension allows to extend a set defined in low-dimensional domain into a higher-dimensional domain. Consider a subset $A \subseteq \Theta$, the cylindrical extension of A to $\Theta \times \Omega$ is denoted $A^{\uparrow \Theta\Omega}$. It is obtained as:

$$A^{\uparrow \Theta\Omega} = A \times \Omega \tag{1.32}$$

Example 1.3. *Let us consider* $\Theta = \{\theta_1,\theta_2\}$. *We want to define* θ_1 *into the two-dimensional space* $\Theta \times \Omega$ *where* $\Omega = \{\omega_1,\omega_2\}$. *Its cylindrical extension is computed as:*
$\theta_1^{\uparrow \Theta\Omega} = \{(\theta_1,\omega_1),(\theta_1,\omega_2)\}$

Projection

Projection is the opposite operation of cylindrical extension. It allows to reduce a set defined in a multi-dimensional domain to a set defined in a lower-dimensional domain. Let C be a subset of $\Theta \times \Omega$. Projecting C on Ω, denoted $C^{\downarrow \Omega}$, means dropping extra coordinates. It is obtained by:

$$C^{\downarrow \Omega} = \{\omega, \omega \in \Omega, C \cap \omega^{\uparrow \Omega\Theta} \neq \emptyset\} \tag{1.33}$$

Example 1.4. *Let us consider the set* $\{(\theta_1, \omega_1), (\theta_2, \omega_1)\}$ *defined on* $\Theta \times \Omega$. *The projection of this set into* Ω *is equal to:* $\{(\theta_1, \omega_1), (\theta_2, \omega_1)\}^{\downarrow \Omega} = \omega_1$.

1.4.3 Vacuous extension and marginalization

Vacuous extension

This operation is useful, when the referential is changed by adding new variables. Thus, a marginal mass function m^{Θ} defined on Θ will be expressed in the frame $\Theta \times \Omega$ as follows.

$$m^{\Theta\uparrow\Theta\Omega}(C) = \begin{cases} m^{\Theta}(A) & \text{if } C = A \times \Omega \\ 0 & \text{otherwise} \end{cases}, A \subseteq \Theta, C \subseteq \Theta \times \Omega. \tag{1.34}$$

It corresponds to make a cylindrical extension of A to $\Theta \times \Omega$. It is the *least committed bba* defined on $\Theta \times \Omega$.

Example 1.5. *Given the following bba defined on* $\Theta = \{Mary, John, Peter\}$ *as:* $m^{\Theta}(\{Mary\})$*=0.5,* $m^{\Theta}(\{John\})$*=0.2,* $m^{\Theta}(\{\Theta\})$*=0.3.*

Let us denote by Ω *the height of the murder where* $\Omega = \{Tall, Short\}$.

The bba defined on Θ *will be defined in a finer frame* $\Theta \times \Omega$ *using the vacuous extension. It is done by taking into consideration all the values of* Ω *for a given value of* Θ *as follows:*

$m^{\Theta\uparrow\Theta\Omega}(\{(Mary, Tall), (Mary, Short)\})$*=0.5,*

$m^{\Theta\uparrow\Theta\Omega}(\{(John, Tall), (John, Short)\})$*=0.2,*

$m^{\Theta\uparrow\Theta\Omega}(\Theta \times \Omega)$*=0.3.*

Marginalization

Given a mass distribution defined on the product space $\Theta \times \Omega$, marginalization corresponds to mapping over a subset of the product space by dropping the extra coordinates. The new belief defined on Θ, $m^{\Theta\Omega\downarrow\Theta}$, is obtained by:

$$m^{\Theta}(A) = \sum_{C \subseteq \Theta\times\Omega, C^{\downarrow\Theta}=A} m^{\Theta\Omega}(C), A \subseteq \Theta \tag{1.35}$$

It corresponds to projecting C on Θ.

Example 1.6. *Let us consider the bba defined on* $\Theta\times\Omega$*:*

$m^{\Theta\Omega}(\{(Mary, Tall), (Mary, Short)\})$*=0.5,*

$m^{\Theta\Omega}(\{(John, Tall), (John, Short)\})$*=0.2,*

$m^{\Theta\Omega}(\{(Mary, Tall)\})$*=0.3.*

Marginalizing $m^{\Theta\Omega}$ *on the coarser frame* Θ*,* $m^{\Theta\Omega\downarrow\Theta}$ *will lead to the following distribution:* $m^{\Theta\Omega\downarrow\Theta}(\{Mary\})$*=0.5+0.3=0.8,* $m^{\Theta\Omega\downarrow\Theta}(\{John\})$*=0.2.*

Note that the vacuous extension of the marginalized bba does not allow to find the initial distribution.

1.4.4 Ballooning extension

This operation is useful if an agent after conditioning realizes that the evidence he has considered as true was not and accordingly he would reconstruct the initial distribution. It can also be useful if beliefs are defined on a limited set and other alternatives were discovered afterwards. The agent should redistribute his beliefs to take them into account.

Let $m^{\Theta}(A|\omega)$ be defined on Θ for $\omega \in \Omega$. To get rid of conditioning, we have to compute its ballooning extension (or conditional embedding (Shafer, 1982)). It is the *least committed* (the least informative) *bba* defined on $\Theta \times \Omega$.

Conditional masses are transferred to C, the largest subset of $\Theta \times \Omega$ whose intersection with the vacuous extension of ω followed by a projection on Θ gives A: $(C \cap \omega^{\uparrow\Theta\Omega})^{\downarrow\Theta} = A$. Thus, $C = (A \times \{\omega\} \cup \Theta \times \overline{\omega})$ where $\overline{\omega}$ stands for the complement of ω according to Ω.

Accordingly, the ballooning extension is defined as:

$$m_{\omega}^{\Theta\hat{\uparrow}\Theta\Omega}(C) = \begin{cases} m^{\Theta}(A|\omega) & \text{if } C = (A \times \{\omega\} \cup \Theta \times \overline{\omega}) \\ 0 & \text{otherwise} \end{cases} \tag{1.36}$$

Example 1.7. *Consider a conditional bba defined on* Θ=$\{Mary, John, Peter\}$ *in context of* $\Omega = \{Tall, Short\}$. *Its definition on* $\Theta \times \Omega$ *is obtained by the application of the ballooning extension. Let us assume that the witness has declared that the murderer is* $Tall$. *The confidence that the murderer is Peter is defined with* $m^{\Theta}(Peter|Tall)$=0.7. *Its corresponding basic belief mass on* $\Theta \times \Omega$ *is obtained by taking into consideration* $\{(Peter, Tall)\}$ *and all the instances of* Θ *for the complement of* $Tall$ *(here* $Short$*). Thus, the mass initially allocated to Peter, given that* $\{Tall\}$ *is a certain event, will be entirely transferred to* {*(Peter,Tall),(Mary,Short),(Peter,Short), (John,Short)*}.
Hence, $m^{\Theta \uparrow \Theta\Omega}$({*(Peter,Tall),(Mary,Short),(Peter,Short), (John,Short)*}) = 0.7.

1.5 Pignistic transformation

Reasoning process is used to know which hypothesis is true for a given problem. So the system, after having analyzed all the available data, should give us the best action to choose. In probability theory, we choose the hypothesis with the highest probability.

Within the TBM, when a decision must be made beliefs held at the credal level induce at a final stage a probability measure. This pignistic measure denoted by $BetP$ (Smets, 1988b) is used to make decisions. It is computed as follows:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \text{ for all A} \in \Theta \tag{1.37}$$

Note that $m(B)$ should be a normalized *bba*.

Example 1.8. *Let us consider that at the credal level we have the following bba:*
$m(\{Mary, John\}) = 0.2$;
$m(\{Peter, John\}) = 0.5$;
$m(\Theta) = 0.3$.
In order to make a decision, we have to compute the pignistic probability $BetP$, *we get:*
$BetP(\{Mary\}) = 0{,}2$;
$BetP(\{John\}) = 0{,}45$;
$BetP(\{Peter\}) = 0{,}35$;
It is more probable that the murderer is John.

1.6 Graphical representation of knowledge

Graphical models (e.g., probabilistic Bayesian networks (Darwiche, 2009; Jensen & Nielsen, 2007; Pearl, 1988), possibilistic networks (Ben Amor et al., 2003; Benferhat & Smaoui, 2007a), credal networks (Cozman, 2000), valuation networks (Shenoy, 1989), belief function networks (Ben Yaghlane & Mellouli, 2008; Xu & Smets, 1996)) are compact representations of uncertainty distributions. Their success is due to their simplicity and their capacity of handling independence relationships. In this section, we briefly recall these networks.

1.6.1 Notations and definitions

Let $V = \{A_1, A_2, \ldots, A_n\}$ be a finite set of variables.

Each variable A_i is associated with a finite set namely its frame of discernment Θ_{A_i} representing all its possible instances.

A graph $\mathcal{G}= (V, E)$ is said to be a directed graph if V is a set of nodes denoting the domain variables and E is a subset from the cartesian product $V \times V$ corresponding to set of directed edges encoding the dependencies among variables.

Some elements of graph theory that have been used in this thesis are recalled here:

- a node A_i is the **parent** of A_j and the node A_j is the **child** of A_i, if there is an arc from A_i pointing towards A_j. $PA(A_i)$ denotes the **parents** of A_i and an instance from the set of parents of A_i (i.e., $PA(A_i)$) is denoted by $Pa(A_i)$. It is defined as an element of the cartesian product of the parents of A_i: $Pa(A_i) = \times\Theta_{A_j \in PA(A_i)}$.
- a **root** is a node with no parents, e.g., A_i is a root node if $PA(A_i) = \emptyset$.
- a **path** is a sequence of nodes from one node to another using the arcs, a path from A_i to A_j is denoted $A_i \mapsto A_j$.
- a **cycle** is a path where the first and the last node are confused into one node, i.e., $A_i = A_j$.
- a **loop** is an undirected cycle.
- a graph is said to be **connected** if it exists a path between each pair of nodes.
- a **DAG** is a *D*irected *A*cyclic *G*raph. When the *DAG* does not contain loops, it is called singly connected. If it contains loops, it is a multiply connected *DAG*.
- an **hypergraph** is a generalization of a graph. An edge can connect any number of nodes.

1.6.2 Brief refresher on Bayesian networks

Probability theory

In probability theory, the static component consists of the assessment of a probability density p on the elements of the frame of discernment Θ such that p: $\Theta \rightarrow [0,1]$:

$$\sum_{A \in \Theta} p(A) = 1 \tag{1.38}$$

Degrees of belief on subsets of Θ are quantified by a probability distribution P such that:

$$\begin{array}{ll} - & \forall \theta \in \Theta, P(\{\theta\}) = p(\theta), \\ - & \forall A, B \subseteq \Theta \text{ with } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B) \\ - & P(A) = \sum_{\theta \in A} p(\theta) \end{array} \tag{1.39}$$

The only dynamic component is the conditioning rule. If $B \subseteq \Theta$ is true and $P(B) \neq 0$, P is updated into the conditional probability distribution $P(.|B)$ such that:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.40}$$

Bayesian networks

A Bayesian network is an associational network. It is defined in two levels:

- A graphical level: a directed acyclic graph (DAG) $\mathcal{G}$ representing the (in)dependence relations in the studied system.
- A numerical level: a quantitative information by associating for each node (i.e., a random variable) a conditional probability that quantifies the effects of its parents on it. The a priori distribution is defined for each node A_i in the context of its parents ($PA(A_i)$) as follows:
  - if A_i is a root node, then the a priori probability of A_i should satisfy:
$$\sum_{a_i} P(a_i) = 1, \text{ where } a_i \in \Theta_{A_i}$$
  - if $PA(A_i) \neq \emptyset$, then the conditional probability over A_i is defined as:
$$\sum_{a_i} P(a_i \mid Pa(A_i)) = 1, \text{ where } Pa(A_i) \in \times\Theta_{A_j, A_j \in PA(A_i)}$$

Computation of the global joint distribution

A Bayesian network satisfies the Markov assumption that agrees with the independencies represented by the network structure. It involves that each variable is conditionally independent from its non-descendants given its parents. Accordingly, the global joint probability distribution over the set $V = \{A_1, \ldots, A_n\}$ is unique and can be expressed as a *product* of the initial conditional probabilities using Bayes rule. The probabilistic chain rule is defined as follows:

$$p(A_1, \ldots, A_n) = \prod_{i=1}^{n} P(A_i \mid Pa(A_i)) \tag{1.41}$$

A posteriori distributions can be recovered from the global joint distribution by marginalization.

Example 1.9. *Let us consider the Bayesian network represented in Figure 1.3.* A_1 *and* A_2 *are root nodes, an a priori probability is specified for them. For the other node, a conditional probability distribution is given, namely* $P(A_3 \mid A_1, A_2)$ *and* $P(A_4 \mid A_3)$.

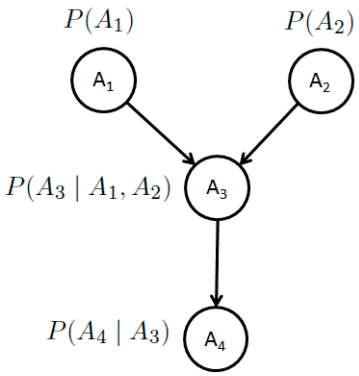


Figure 1.3: A Bayesian network

The global joint distribution is computed for these local distributions using the chain rule (Equation 1.41) as follows:

$$p(A_1, A_2, A_3, A_4) = P(A_1) \cdot P(A_2) \cdot P(A_3 \mid A_1, A_2) \cdot P(A_4 \mid A_3)$$

1.6.3 Brief refresher on possibilistic networks

Possibility theory

The possibility theory was introduced by Zadeh (1978) and developed by Dubois and Prade (1982, 1988, 1998).

A possibility distribution π is a mapping from the set of interpretations Θ to the unit interval [0,1]. To this scale, two interpretations can be attributed, a *quantitative* one when values have a real sense and a *qualitative* one when values reflect only an order between the different states of the world.

It represents a state of knowledge about a set of possible interpretations distinguishing what is plausible from what is less plausible. The value $\pi(\theta)$ expresses a degree of coherence of the interpretation θ with respect to available knowledge encoded by π. By convention, $\pi(\theta) = 0$ means that the interpretation θ is impossible, and $\pi(\theta) = 1$ means that nothing prevents θ from being the real world.

When $\pi(\theta_1) > \pi(\theta_2)$, θ_1 is a preferred candidate to θ_2 for being the real state of the world. π is thus a convenient encoding of a preference relation that can embody concepts such as plausibility or consistency.

Given a possibility distribution π, we can define a mapping grading the possibility measure of an event $A \subseteq \Theta$ as follows:

$$\Pi(A) = \max_{\theta \in A} \pi(\theta). \tag{1.42}$$

The dual measure of Π is the necessity measure defined as:

$$N(A) = 1 - \Pi(\overline{A}) \tag{1.43}$$

Conditioning is consists in modifying the initial knowledge, encoded by a possibility distribution π, by the arrival of a new piece of information $A \subseteq \Theta$. We assume that A is not empty and that $\Pi(A) > 0$. There are two main definitions of possibilistic conditioning.

The definition proposed for the qualitative setting is called *min-based conditioning* and is defined by:

$$\pi(\theta|A) = \begin{cases} 1 & \text{if } \theta \in A, \Pi(A) = \Pi(\theta), \\ \pi(A) & \text{if } \theta \in A, \Pi(A) > \Pi(\theta), \\ 0 & \text{if } \theta \notin A \end{cases} \tag{1.44}$$

The definition proposed for the quantitative setting is called *product-based conditioning* and is defined by:

$$\pi(\theta|A) = \begin{cases} \dfrac{\pi(\theta)}{\Pi(A)} & \text{if } \theta \in A \\ 0 & \text{otherwise} \end{cases} \tag{1.45}$$

The two definitions satisfy the equation of the form:

$$\pi(\theta) = \pi(\theta|A) \diamond \Pi(A), \tag{1.46}$$

It is similar to Bayesian conditioning, $\diamond$ is either min or product. The rule based on the product is much closer to genuine Bayesian conditioning than the qualitative conditioning defined from the minimum which is purely based on the comparison of levels; product-based conditioning requires more of the structure of the unit interval.

Possibilistic networks

A possibilistic network is defined in two levels:

- A graphical level: a Directed Acyclic Graph (DAG) $\mathcal{G}$ representing the (in)dependence relations in the studied system.
- A numerical level: a quantification of different links of the DAG by conditional possibilities of each node A_i in the context of its parents $PA(A_i)$. Conditional distributions should respect the following constraints:
  - if $PA(A_i) = \emptyset$, then the a priori possibility distribution relative to A_i should satisfy:

    $$max_{a_i}\pi(a_i) = 1, a_i \in \Theta_{A_i}$$

  - if $PA(A_i) \neq \emptyset$, then the conditional possibility distribution relative to A_i in the context of its parents should satisfy:

    $$max_{a_i}\pi(a_i|Pa(A_i)) = 1, a_i \in \Theta_{A_i}, Pa(A_i) \in \times\Theta_{A_j, A_j \in PA(A_i)}$$

Computation of the global joint distribution

Since in conditioning, there is a product-based (resp. min-based) conditioning for the quantitative (resp. qualitative) setting, possibilistic networks can also be divided into quantitative vs. qualitative networks. Accordingly, the computation of the global joint distribution differs according of the kind of the possibilistic network.

Quantitative possibilistic networks On a quantitative possibilistic network, we can compute the joint possibility distribution on the set of variables $V = A_1, A_2, \ldots, A_n$ using the following equation:

$$\pi(A_1, A_2, \ldots, A_n) = \sum_{i=1}^{n} \pi(A_i|Pa(A_i)) \tag{1.47}$$

Qualitative possibilistic networks When only the ordering induced from the possibility degrees is important and not the values themselves, qualitative possibilistic networks are used to describe the system.

Their corresponding joint possibility distribution are computed as follows:

$$\pi(A_1, A_2, \ldots, A_n) = \min_{i=1}^{n} \pi(A_i|Pa(A_i)) \tag{1.48}$$

1.6.4 Brief refresher on credal networks

The extension of Bayesian networks to deal with imprecision in probability is achieved by means of the notion of credal sets which are closed convex sets of probability mass functions. A credal set for a random variable A_i is denoted by $K(A_i)$. The network where each conditional probability function is replaced by a conditional credal set is called a credal network.

Credal Networks (Cozman, 2000) are imprecise probabilistic models as advocated by Walley (Walley, 1991). They provide a representation for imprecise probabilistic knowledge through direct acyclic graphs (DAGs). They are also used to evaluate the robustness of Bayesian networks (Cozman, 1997).

Accordingly, a credal network may be viewed as a Bayesian network where some (or all) parameters are not precisely known, but instead constrained by convex constraints.

Example 1.10. *Let us consider the network in Figure 1.3. For instance, constraints may be:*

- $P(a_1) \in [0.1, 0.3]$
- $P(a_3|a_1, a_2) = 0.5$
- $P(a_4|a_3) + P(a_4|\overline{a_3}) \leq 0.75$

In a credal network every node is associated with a variable, and every variable is associated with a collection of local credal sets denoted by $K(A_i \mid pa(A_i))$, where $pa(X)$ denotes the parents of variable A_i in the graph. Thus, a node stores the credal sets such that:

$$\{K(A_i \mid pa(A_i) = \pi_1), \ldots, K(X \mid pa(A_i) = \pi_m)\}, \tag{1.49}$$

where $\{\pi_1, \ldots, \pi_m\}$ are the instances of $pa(A_i)$. A root node has only one credal set associated with it.

Computation of the global global joint distribution

In the literature, the most adopted concept for credal networks is that every variable is strongly independent of its non-descendants given its parents. In this case, a credal network can be viewed as a collection of Bayesian networks that share the same graph, the global distribution is defined in a similar way to Bayesian networks:

$$p(\mathcal{X}) = \prod_{i=1}^{n} P(A_i \mid pa(A_i)), \mathcal{X} = \{A_1, \ldots, A_n\} \tag{1.50}$$

1.6.5 Valuation networks

Valuation based-system

A valuation-based system (VBS) is a formal mathematical system for representation of knowledge and reasoning with it (Shafer, 1992; Shenoy, 1993, 1994). It is an abstract framework that allows to uniformly represent probability theory, belief-function theory as well as possibility theory. It has some similarities to influence diagrams. However, unlike influence diagrams which emphasize conditional independence among random variables, valuation-based systems

emphasize factorizations of joint probability distributions. Also, whereas influence diagram representation allows only conditional probabilities, valuation-based system representation allows all probabilities.

The representation of knowledge is made with entities called variables and valuations. A valuation is an object that encodes knowledge about variables in a subset. It can be combined, marginalized, and removed over the set of all valuation denoted V as follows:

- A *combination* is a mapping $\oplus : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$, such that x is a valuation on Θ_X and y is a valuation on Θ_Y, then $x \oplus y$ is a valuation of $\Theta_X \cup \Theta_Y$.
- A *marginalization* is a mapping $\downarrow (\Theta_X \setminus \{\Theta_A\}) : \mathcal{V}_{\Theta_X} \rightarrow \mathcal{V}_{\Theta_X \setminus \{\Theta_A\}}$, such that if x are valuations Θ_X, then $x^{\downarrow(\Theta_X \setminus \{\Theta_A\})}$ is a valuation for $(\Theta_X \setminus \{\Theta_A\})$
- *removal* is a mapping $\ominus : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$, It can be regarded as an inverse of combination (Shenoy, 1994) such that if x is a valuation on Θ_X and y is a valuation on Θ_Y, then $x \ominus y$ is a valuation of $\Theta_X \cup \Theta_Y$

The valuation based system is able to represent several uncertain frameworks by translating valuations and operations to their special interpretation in the corresponding theory.

Within belief function theory, a valuation defined on Θ_X corresponds to the basic belief assignment assigned to subsets of Θ_X. Dempster's rule of combination, the marginalization operation and the decombination operation in belief function theory corresponds to respectively the combination operation, the marginalization operation and the removal operation on VBS.

Valuation networks with belief functions

A graphical representation of a VBS is called a valuation network. When uncertain knowledge is formalized with belief functions, the valuation network is called an evidential network. It consists of an hypergraph as shown in Figure 1.4 that can be defined in two components (Shenoy, 2000):

- **Graphical component:**
  - Variables:
    - *decision nodes*: depicted with rectangles.
    - *chance nodes*: depicted with circles.
  - Valuations:
    - *indicator nodes*: depicted with double-triangles. They represented qualitative constraints on the joint frames of the variables.
    - *utility nodes*: depicted with diamonds. They represent factors of the joint utility.
    - *probability nodes*: depicted with triangles. They represent multiplicative factors of the family of joint probability distributions of the chance variables in the problem.
  - Links
    - *edges*: undirected links connecting variables to potentials and utility functions.
    - *arcs*: directed links between variables. They define the information constraints.
- **Numerical component:** representing our knowledge by valuations. Valuations are interpreted as belief functions and are linked to the variables in their domain.

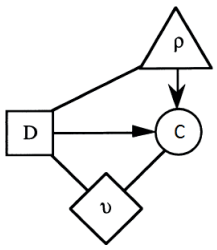


Figure 1.4: A valuation network

Valuation networks with conditional beliefs

Cano et al. (1993) and Shenoy (1993) have presented an axiomatic system for propagating uncertainty including belief functions in Bayesian networks, based on Shafer and Shenoy's axiomatic framework (Shenoy & Shafer, 1990). However, the belief functions for representing relations of the variables in their system are still represented on *the product space.*

To avoid confusion, with the classic meaning of conditional beliefs used in Bayesian networks, in (Xu & Smets, 1996), the authors change the term of conditional beliefs as *non-informative* belief functions and define them as:

Definition 1.1. *Given two disjoint subsets X and Y and let bel be a belief function defined on the* ***joint space*** $\Theta_X \times \Theta_Y$*. It is said that bel is a conditional belief function for Y given X if and only if $bel(Y \mid \Theta_X)$ is a vacuous belief function over X.*

The term "non-informative" was chosen to define this belief function since it gives some information about variables in Y and their relationship with variables in X, but no information about X.

Computation of the global joint distribution

The global joint distribution is computed by combining local joint distributions (m_i) as follows:

$$m = \oplus\{m_i \mid i = 1 \ldots n\} \tag{1.51}$$

Example 1.11. *Let us consider the network presented in Figure 1.5. It is a valuation networks composed of four variables (W, X, Y and Z) and four valuations (m_1 for $\{W\}$, m_2 for $\{W, X\}$, m_3 for $\{X, Y\}$ and m_4 for $\{Y, Z\}$).*

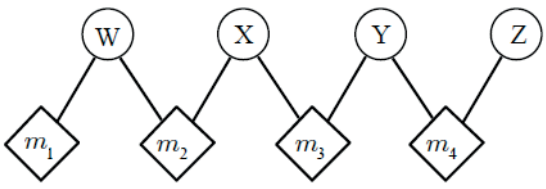


Figure 1.5: A valuation network with conditional beliefs

The global joint distribution is computed using Equation 1.51 as follows:

$$m = m_1 \oplus m_2 \oplus m_3 \oplus m_4$$

Joint beliefs vs conditional beliefs

Let $V = \{A_1, A_2, \ldots, A_n\}$ be a set of variables. Associated with A_i, there is a frame Θ_{A_i} which is the set of all its possible values. Let X and Y be two disjoint subset of V. Their frames, Θ_X and Θ_Y, are the product spaces of the frames of the variables they include.

A conditional belief function for Y given X is represented by $bel^Y(. \mid x)$ where $x \in \Theta_X$, which means that we know the belief about Y given that we only know that the actual value of X is in x. Joint belief functions on X and Y are defined on the space $\Theta_X \times \Theta_Y$.

Expressing beliefs in terms of conditional distributions is more natural and easy for the users to provide and to understand. In fact, given two disjoint subsets $X, Y \subseteq V$, to represent conditional belief functions for Y given X by a joint form, one needs $2^{\Theta_X \times \Theta_Y}$ elements in the worst case, while to represent them by conditional form one only needs $2^{\Theta_X + \Theta_Y}$ elements in the worst case.

Example 1.12. *Given two variables A_1 and A_2 where $\Theta_{A_1} = \{a_{11}, a_{12}\}$ and $\Theta_{A_2} = \{a_{21}, a_{22}\}$. Let us consider the assertion: "if $A_1 = a_{11}$ then $A_2 = a_{22}$ with a bbm of 0.8".*

This relation is represented on the belief joint space $\Theta = \Theta_A \times \Theta_B$ $= \{(a_{11}, a_{21}), (a_{11}, a_{22}), (a_{12}, a_{21}), (a_{12}, a_{22})\}$ as:
- $m(\{(a_{11}, a_{22}), (a_{11}, a_{21}), (a_{12}, a_{21})\}) = 0.8$
- $m(\Theta) = 0.2$.

With conditional distributions, it is simply defined as:
- $m(a_{22} \mid a_{11}) = 0.8$
- $m(\Theta_{A_2} \mid a_{11}) = 0.2$
- $m(\Theta_{A_2} \mid a_{12}) = 1$
- $m(\Theta_{A_2} \mid \Theta_{A_1}) = 1$

Even though not all belief functions on $\Theta_X \times \Theta_Y$ admit an equivalent representation by a set of conditional belief functions, it is assumed that the experts' knowledge is encoded in the conditional form and that the joint beliefs they would provide are those based on the known conditional form.

1.6.6 Directed belief networks with conditional beliefs

There exist several associational belief networks. We can categorize them according to the way conditional distributions are defined: those where conditional distributions are defined for all parents as for Bayesian networks and those where conditionals are defined per single parent (per edge). Note that, under the belief function theory it is possible to avoid choosing the appropriate a priori since there is no distinction between the De re and the Dicto relations (Smets, 1992) whereas in probabilistic models doing such a way leads to the Lewis trivialization (Lewis, 1976).

In this subsection, we present a brief recalling of belief networks. A detailed analysis is carried and exposed in Chapter 3.

Conditional distribution for all parents

Definition Represented as a Bayesian network, a belief function network (Simon et al., 2008) is represented with a directed acyclic graph (DAG), denoted by $\mathcal{G}$ called *Evidential Network (EN)*. It combines the belief function theory with a DAG.

An Evidential network is defined on two levels:

- Qualitative level: represented by a DAG, $\mathcal{G} = (V, E)$ where edges encode the dependencies among variables. Each variable A_i is associated with a finite set namely its frame of discernment Θ_{A_i} representing all its possible instances. An instance from the set of parents of A_i (i.e., $PA(A_i)$) is denoted by $Pa(A_i)$.
- Quantitative level: represented by the set of *bba*s associated to each node in the graph.
  - For each root node A_i (i.e., node without parent nodes) having a frame of discernment Θ_{A_i}, an a priori mass distribution m^{A_i} is defined on the powerset $2^{\Theta_{A_i}}$.
  - For other nodes, a conditional *bba* $m^{A_i}(.|Pa(A_i))$ is specified for each value of A_i knowing the value of all the parents $Pa(A_i)$. It means that a conditional belief mass table represents the relation between the basic belief masses expressed on the frame of discernment of the child node and the basic belief masses defined on the frame of discernment of the variables in the parent nodes.

The evidential network is proposed in context of reliability modeling where the domain associated with variables is defined as {Up, Down}. Its corresponding power set is {$\emptyset$, {Up}, {Down}, {Up, Down}}.

Computation of the global joint distribution In Evidential Networks, the computation of inference is made with the total probability theorem or the Bayes theorem extended to belief masses. Therefore, the joint belief chain rule is computed by multiplying all conditional beliefs as follows:

$$m^V = \prod_{i=1}^{n} m^{A_i}(. \mid Pa(A_i)) \tag{1.52}$$

Propagation in ENs is based on the Bayesian model. In fact, exact algorithms used for inference in Bayesian networks, e.g,. the junction tree (Jensen et al., 1990), are directly used.

Conditional distribution per single parent

Definition Even though graphically represented as Bayesian networks, Evidential Networks with Conditional beliefs (ENC) (Smets, 1993b; Xu & Smets, 1996) and their generalization Directed EVidential Networks ($DEVN$) (Ben Yaghlane & Mellouli, 2008) define conditional beliefs in a different way from that one traditionally used in Bayesian networks and Evidential networks. These networks can be defined on two levels as follows:

- Qualitative level: represented by a directed acyclic graph (DAG), $\mathcal{G} = (V, E)$. Each variable A_i is associated with a finite set namely its frame of discernment Θ_{A_i} representing all its possible instances. If there is an edge from variable A_i to variable A_j, A_i is called a parent of A_j. Parents of A_i are denoted by $PA(A_i)$ and a single parent is denoted by $PA_j(A_i)$

- Quantitative level: represented by the set of beliefs (Bel) associated to each *edge* in the graph. An edge represents a conditional relation between the two nodes it connects.

The number of conditional relations in the DEVN and ENC depends on the *number of edges* whereas in ENs it depends on the *number of child nodes* (Laâmari et al., 2010) which simplify knowledge acquisition. Figure 1.6 shows a network with individual causes.

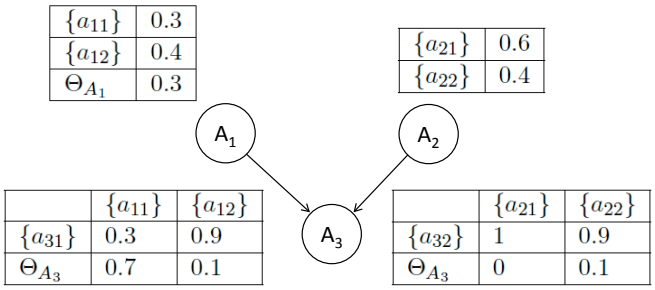


Figure 1.6: The directed evidential network

Notice that the knowledge about the relations between two nodes in the belief function network can be issued from different sources (local conditional beliefs). The knowledge is then aggregated by computing the global belief from all local conditional beliefs.

Computation of the global joint distribution The definition of the joint distribution under a belief function framework is different from the construction made in Bayesian networks. In fact, it is obtained by combining the joint distribution of each node. In (Ben Yaghlane et al., 2003), the authors present the following belief chain rule respecting the independence principle (Ben Yaghlane et al., 2002b, 2002a):

$$bel^{A_1,\ldots,A_n} = \bigcirc\!\!\!\!\!\!\cap_{i=1,\ldots,n} \Big(\bigcirc\!\!\!\!\!\!\cap_{\omega \in PA(A_i)} bel^{A_i}(.|\omega)^{\uparrow A_i \times PA(A_i)} \Big) \tag{1.53}$$

1.7 Conclusion

In this chapter, we presented the basic concepts of the belief function theory to represent and reason about different kinds of imperfect knowledge. We saw that belief function graphical models can be either undirected (Valuation Networks) or directed graphs (Evidential Networks with Conditional beliefs, Directed EVidential Networks, Evidential Networks). We explained that directed graphs are more appropriate for representing conditional belief functions. In fact, it is easier to collect and to express beliefs in terms of conditionals instead of joint belief functions.

The belief function theory will be used as a tool to formalize the imperfect causal knowledge which may be represented with a causal belief network. For this reason, we will focus on the notion of causality in the next chapter.

Chapter 2

Causality modeling

2.1 Introduction

Causality is an important concept involved in a wide range of applications such as genetics, physics, psychology as well as social science. The notion of causality was introduced since more than 200 years and has long been recognized as an important research axis. It plays a crucial role in the expression of our perception of our environment. In fact, it enables to anticipate the dynamics of events when the system is evolving and thus choose the right actions to achieve the goals. Accordingly, discovering causal relations is a task of crucial importance.

The idea of causality is undefinable if a general and precise definition is sought (not restrained to particular cases) (Zadeh, 2001). In fact, it can be seen as a regular association (Hume, 2006), a counterfactual dependence (Lewis, 2004) or a probability raising (Eells, 1991). The concept of causality is connected to other ideas like those of explanation or responsibility what makes difficult its comprehension and its modeling.

In the Artificial Intelligence field, rather recently researchers were interested in the problems arising from the modeling of causality (e.g., Shafer (1996), Pearl (1998, 2000) and Halpern and Pearl (2005)). In fact, it is important to provide the systems of inference or decision-making with explanations capacity for an operator or human user (Dubois & Prade, 2003).

Identifying causal relations differs from diagnosing. Indeed, standard diagnosis problems consist of making inspections to find the history of the faults that explain the observations and therefore provide an explanation for what components failed and when they did (Otero & Otero, 2000). Ascribing causes (Benferhat et al., 2008; Bonnefon et al., 2008) is a different problem: it consists of identifying what elements in a set of observed or intervened events in a sequence are causally related, on the basis of some background knowledge about the normal course of things. Causality is therefore considered as a nonmonotonic relation.

To achieve this goal, one may use observational data that provide some information about the statistical relations among events. However, we should be aware since associations do not necessarily follow a causal process. To tackle this problem, we use interventional data

(Pearl, 2000). An intervention is an external action that perturbs the spontaneous behavior of the system by forcing a variable to take a specific value. Through these experimentations, the effects of all direct (and undirected) causes related to the variable of interest will be ignored. Therefore, given two dependent variables A_1 and A_2, if an action on an event A_1 has no impact on an event A_2, then A_1 cannot be the cause of event A_2, but if a manipulation on the event A_1 leads to a change in A_2, then we can conclude that A_1 is a cause of A_2.

In the context of observations any representation of the background knowledge is suitable whereas in the context of interventions the graphical structure is needed. In fact, graphical models allows to delimit the mechanisms affected by such disturbances. Interventions will be represented on this causal structure by the mean of the "do" operator. This tool was originally introduced by (Goldszmidt & Pearl, 1992) for the ordinal conditional functions of Spohn (1988) and proposed after that in (Pearl, 2000) for causal Bayesian networks.

In this chapter, we first remind some definitions of causality in Section 2.2. The distinction between interventions and observations is detailed in Section 2.3. In Section 2.4 functional causal models are introduced. Section 2.5 is consecrated to another artificial intelligence based model of causality namely event trees. Section 2.6 is dedicated to a stochastic ascription and modeling of causality. Nonmonotonic approaches for ascribing causal relations are presented in Section 2.7.

2.2 Definitions

2.2.1 Causation vs association

Causation and association are different ideas even though they are related. In fact, statistical association by itself does not prove causation (Scheines, 2008). However, causal relation can produce automatically statistical correlation. Two events are highly associated if we observe one of them, we are in good position to say that the other one is observed. An exterior intervention on one of them will not change the other one.

The fundamental property of association is *symmetry*. In other words, if learning that the occurrence of a first event A_1 changes your certainty about another event A_2, then learning about A_2 is also informative about A_1 (i.e., A_1 associated to A_2 entails that A_2 is also associated to A_1). However, causation is an *asymmetric* relation (i.e., A_1 caused A_2 does not imply that A_2 caused A_1).

However, it does not mean that symmetric cases of causation are prohibited. It only means they are not necessary (e.g., losing sleep can cause anxiety, and anxiety can also cause a loss of sleep).

Example 2.1. *Let us denote by S, the fact of smoking and by A the fact of drinking alcohol. The correlation coefficient between S and A is almost certainly quite positive. Moreover, it is a symmetric relation. Does this prove that smoking causes drinking alcohol? Obviously, no.*

Now, consider the case of lung cancer denoted by C. Let us investigate the relation between C and S. These two events are correlated and their association involves causation. It is an

asymmetric relation. In fact, smoking causes lung cancer but lung cancer does not cause smoking.

2.2.2 Genuine vs spurious causes

A *genuine* cause should be well distinguished from a *spurious* cause (Halpern & Pearl, 2005). We call genuine a factor correlated with an effect (or is associated with it) and is judged to be causal a genuine cause and a factor that is correlated with an effect (or is associated with it) but is judged to be noncausal a spurious cause (Suppes, 1970).

Spurious causes may be classified into two categories. The regularity between some spurious causes and the effect can be plausibly explained by an alternative cause. Other spurious causes are not clearly explained by any alternative cause, but might be correlated with the effect purely by chance.

Example 2.2. *Consider that students in a psychology class who had long hair got higher scores on the midterm than those who had short hair, there would be a correlation between hair length and test scores. Not many people, however, would believe that there was a causal link and that, for example, students who wished to improve their grades should let their hair grow. The real cause might be gender: that is, women (who usually have longer hair) did better on the test. Or that might be a spurious relationship too. The real cause might be class rank: Seniors did better on the test than sophomores and juniors, and, in this class, the women (who also had longer hair) were mostly seniors, whereas the men (with shorter hair) were mostly sophomores and juniors (Vogt, 2005).*

2.2.3 Counterfactuals

Counterfactual reasoning is about what would have happened if events other than the ones currently observed had happened. A counterfactual (contrary-to-fact) is a conditional statement. It carries the suggestion that the antecedent of such a conditional is false.

Hume (2006) and recently Pearl and Hopkins (2007), consider that a causal relation of the form event A caused event B, can be explained in terms of conditionals of the form "If A had not occurred, B would not have occurred".

The meaning and truth of counterfactuals is based on the assumption of the existence of possible worlds (interpretations) and a similarity metric over possible worlds as understood in modal logic. Hence, the meaning of "if A had not happened, then B would not have happened either", is for each possible world W_1 in which A did not happen and B did happen, there is at least one world W_2 in which A did not happen and B did not happen that is closer to the actual world than W_1.

Example 2.3. *Let us consider this claim: "If he had hurried, then he would have caught the bus". The implication is that he has actually not hurried and did not caught the bus. It supposes that among the possible worlds where he had hurried, it exists at least one world where he would have caught the bus. This world will be the most realistic one: closer than the one where he had hurried and has not caught the bus.*

2.3 Observations vs Interventions

2.3.1 Observation

Observing (seeing or the act of watching) is a detailed examination of something when the experimental conditions are static. It can provide some information about the statistical relations amongst events. When we have passively observed an event, we can reason backwards diagnostically to infer the causes of this event, or we can reason forward and predict future effects (e.g., observing someone smoking).

Formally, observations are modeled by setting the event variables to the values that have been observed. Based on the chain rule, the probabilities of other events conditional on the observed variable can be computed, i.e., giving the spontaneous behavior of a variable (Ben Yaghlane et al., 2002a; Vejnarová, 2012). The structure of the causal model is crucial for these computations.

2.3.2 Intervention

Intervening (Pearl, 2000; Spirtes et al., 2001; Woodward, 2003) (doing or the act of manipulating) is the effect of an external action to the system that forces a variable to have a specific value (e.g., forcing someone to take smoking nicotine patch). It means that the natural behavior of an object is voluntary changed.

Interventions allow the identification of elements in a sequence of events that are related in a causal way. In fact, a paradigmatic assertion in causal relations is that the exterior manipulation of a genuine cause will result in the variation of the effect. Interventions are therefore used for causal discovery to arbitrate between causal structures that fit the correlations equally well (see Figure 2.1).

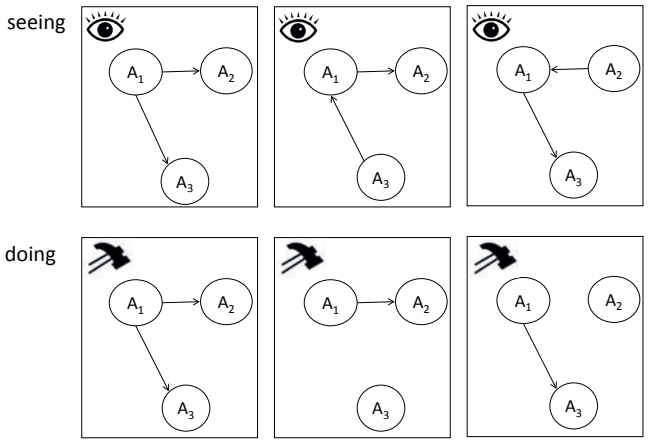


Figure 2.1: Seeing vs doing

Interventions enable to anticipate the dynamics of events when the system is evolving. Accordingly, to compute the effect of interventions, the application of Bayes rule is not ap-

propriate. Tools to compute the effect of interventions, i.e., giving that an action is performed, are therefore needed.

2.4 Functional causal models

Functional causal models express causal relationships in form of deterministic equations and probabilities introduced through unobserved error terms. It consists of a set of equations representing mechanisms of the form of:

$$A_i = \mathrm{f}(Pa_i, U_i) \tag{2.1}$$

The equality has a deterministic meaning. The notation Pa_i stands for the markovian parents of a variable A_i that are the set of variables judged to be immediate causes of A_i. The variable A_i is further called dependent variable of Pa_i. U_i represents errors or disturbances due to omitted factors. A set of equalities in the form of Equation 2.1 is called *Structural Equation Model (SEM)*.

SEM is considered as a mathematical tool for drawing causal conclusions from a combination of observational data and theoretical assumptions (Wright, 1921; Duncan, 1975).

Halpern and Pearl (2005) propose a framework which distinguished a priori *endogenous variables* (whose values are governed by structural equations, corresponding for example to physical laws) and *exogenous variables* (determined by factors external to the model). The definition of causality in this context is closely linked to the idea of counterfactuals. Thus, the fact A, which is a subset of endogenous variables, took specific values, is the real cause of an event A_2, if:

- A_1 and A_2 are true in the real world;
- this subset is minimum;
- another assignment of values to this subset of variables would make A_2 false;
- the values of other endogenous variables not directly involved in the production of A_2 being fixed in a certain way;
- if A_1 alone is sufficient to cause A_2 in this context.

Directed graphs, called *causal diagrams*, are used with SEM to better represent causal relationships (e.g. (Verma & Pearl, 1990)). Hence, an arc in the DAG pointing from PA_i to A_i corresponds to a functional relationship. Each set of child-parent (called family) is a mechanism (an equation) that can be treated separately.

The weakness of this representation, is its lack of selective power since only endogenous variables can be causes or be caused. Besides, incomplete or insufficient information may render the construction of structural equation impossible.

Example 2.4. *Let us consider two variables A_1, A_2 and A_3 and the following system:*

$$A_1 = (A_2 \vee \neg A_3) \wedge U_1$$

A_1 has A_2 and A_3 as causes. U_1 represents disturbances. It can be represented with a graph as depicted in Figure 2.2.

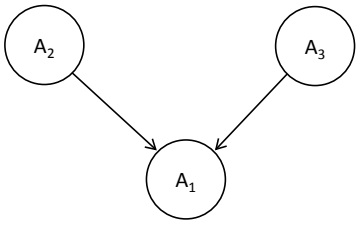


Figure 2.2: A causal diagram

2.5 Event trees

An event tree is a set of objects (situations) partially ordered by time. This approach (Shafer, 1996) is based on *causal logic*. Events are represented on a tree where each node represents a situation encoding an uncertain measure and each node corresponds to a possible choice that may be uncertain.

It allows the representation of the temporal sequence between events interconnected not only by the precedence relations but also relations of specificity and implication. Five basic relations are possible between two events A_1 and A_2. An event A_1 can:

- be a specialization of A_2 or overlaps A_2 (if A_1 happens, A_2 also happens at the same time);
- require for its realization A_2 (if A_1 happens, A_2 has already happened);
- announce A_2 or foretell A_2 (if A_1 happens, A_2 happens after);
- possibly be followed by A_2 or forbear A_2(if A_1 happens, it is possible that A_1 comes after);
- exclude A_2 or diverge from A_2 (if A_1 arrives, A_2 not happen).

Example 2.5. *Let us consider these networks presented in Figure 2.3. The relation between A_1 and A_4 is a relation of precedence in both networks. However, it is only in Figure 2.3 (b) where A_1 foretells A_4. In fact, A_1 foretells A_4 if any path down the tree that passes through A_1 must later passes through A_4.*

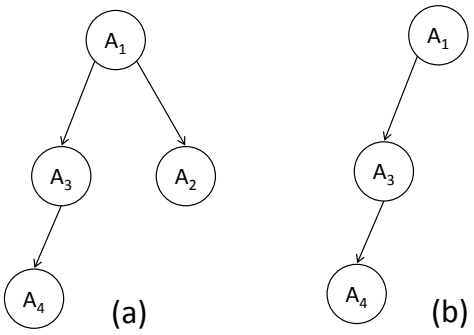


Figure 2.3: Event trees

2.6 Stochastic modeling of causality

Probabilistic causality definitions are given when uncertain information are explicitly described by means of probability distributions. Initially, it is viewed as simple probability changes. Indeed, the idea behind probabilistic causation is that causes change the probabilities of their effects. Using conditional probability, an event A is seen as a cause of an event B if the probability degree of B raises in the light of the new information A (Good, 1961a, 1961b).

Formally, A is a cause of B if the probability that B occurs given that A occurs, is higher than the unconditional probability that B occurs:

$$P(B|A) > P(B) \tag{2.2}$$

For that, one should simply perform a regression on B on its own past, to recover the variance of the residuals and then compare it with the results obtained from the regression of B on its own past and the past of A.The higher the difference between the two variances is, the stronger the events will be causally related (Granger, 1980).

Probabilistic causation is alternatively defined as A causes B if the occurrence of A raises the probability of B (Eells, 1991). Thus, the probability that B occurs, given that A occurs, is higher than the probability that B occurs, given that A does not occur. Formally,

$$P(B \mid A) > P(B \mid \overline{A}) \tag{2.3}$$

However, these definitions are symmetric which is not a desirable property in causality ascription. Indeed, one cannot say if A is a cause of B or B is a cause of A since:

$$P(B \mid A) > P(B \mid \overline{A}), \text{ if and only if } \; P(A \mid B) > P(A \mid \overline{B}) \tag{2.4}$$

Causality should be an asymmetric relation in the sense that if A causes B then B will not also cause A. To satisfy this property, other definitions of causality have been proposed (Mellor, 1995; Suppes, 1970). They implicitly integrate the notion of time in the characterization of causal relation. They stipulate that causes precede their effects in time.

Moreover, initial definitions of causality (i.e., Equation 2.2 and Equation 2.3) have trouble with spurious correlations (Suppes, 1970), i.e, A and B are correlated events and may be

effects of a common cause C. In this case, it may be that $P(B \mid A) > P(B \mid \overline{A})$ even though A does not cause B.

To overcome these limitations, the concept of screening-off was introduced to describe the following situation:

$$\text{If } P(B \mid A\&C) = P(B \mid A), \text{ then } A \text{ is said to screen } C \text{ off from } B \tag{2.5}$$

Accordingly, probabilistic causality is defined as (Suppes, 1970):

Definition 2.1. *An event A'_t occurring at time t' causes an event B_t occurring at time t where t' is later than t if:*

1. $P(B_{t'} \mid A_t) > P(B_{t'} \mid \overline{A}_t)$
2. *There is no further event $C_{t''}$, occurring at a time t'' earlier than or simultaneously with t, that screens $B_{t'}$ off from A_t.*

Note that ascribing causes according to Definition 2.1, do not take into consideration the strength of the causal relation (Salmon, 1980).

2.6.1 Causal networks

Graphical models are important tools proposed for an efficient representation and analysis of uncertain information commonly used by an increasing number of researchers. The success of graphical representations is due to their capacity of representing and handling independence relationships. They allow a local representation and reasoning easily supported by human mind. Causal networks allow to model cause-effect relations.

Indeed, causal Bayesian networks provide formal semantics to the notion of interventions, which plays an important role for eliciting causal relations between variables. In fact, several researchers have emphasized the importance of causal learning from interventions over learning from simply observing data to arbitrate between causal structures that fit the correlations equally well and have shown how passive observations are used to guide interventions toward maximally informative targets (e.g., (Glymour, 2001; Steyvers et al., 2003; Meganck et al., 2006)).

Definition

A causal Bayesian network (Pearl, 2000) is a probabilistic model where edges represent causal relationships. However, a probability distribution, can be represented by several equivalent Bayesian networks if they describe exactly the same conditional independence relation and induce the same joint distributions due to Markov equivalence. Only one of these networks follows the causal process, the so-called causal network.

Accordingly, causal networks provide a convenient framework for causal modeling and reasoning as they have a stricter interpretation of the meaning of edges. A causal Bayesian network, as shown in Figure 2.5, consists of:

- A graphical component: It is a directed acyclic graph (DAG) encoding a set of independence relations, where nodes represent variables and arcs describe *cause-effect* relations.

The causal process follows the direction of the edges. Thus, an event is a cause of its child node and an effect of its parent node. The set of parents of A_i is denoted by $\text{Pa}(A_i)$.

- A numerical component: It is a quantitative information by associating for each node (i.e., a random variable) a conditional probability distribution that quantifies the effects of its parents on it. The a priori distribution is defined for each node A_i in the context of its parents ($PA(A_i)$) as follows:
  - if A_i is a root node, then the a priori probability distribution of A_i should satisfy:

$$\sum_{a_i} P(a_i) = 1, \text{ where } a_i \in \Theta_{A_i}$$

  - if $PA(A_i) \neq \emptyset$, then the conditional probability over A_i is defined as:

$$\sum_{a_i} P(a_i \mid Pa(A_i)) = 1, \text{ where } Pa(A_i) \in \times\Theta_{A_j, A_j \in PA(A_i)}$$

Example 2.6. *Let us consider the networks presented in Figure 2.4. They concern a description of knowledge regarding the link between smoking and having yellow teeth (associational networks). These networks are equivalent in terms of independence relations. Nodes are described as follows:*

- *S describes the fact of smoking, $\Theta_S = \{s_1, s_2\}$ where s_1 is yes and s_2 is no.*
- *T represents having yellow teeth, $\Theta_T = \{t_1, t_2\}$ where t_1 is yellow and t_2 is otherwise.*

From these networks, only one follows the causal process (see Figure 2.5).

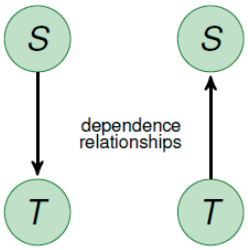


Figure 2.4: Associational equivalent networks

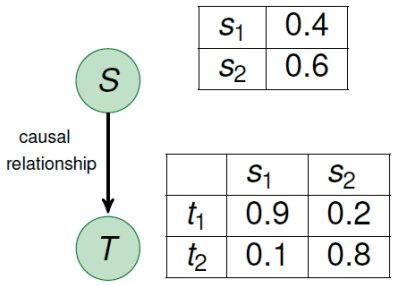


Figure 2.5: A causa Bayesian network

Properties of a causal Bayesian network

A causal Bayesian network has the following properties (Pearl, 2000; Spirtes et al., 2001):

- Causal Markov assumption: a node is independent of its distant causes given its direct causes. Bayes rule is therefore still valid on the causal network.
- Causal sufficiency assumption: there exist no common hidden variables in the domain that are parent of one or more observed variables of the domain.

- Causal faithfulness assumption: a distribution is faithful to the graphical structure $\mathcal{G}$ if and only if exactly the independence facts true in the distribution are entailed by the graphical structure $\mathcal{G}$.

The "do" operator to represent interventions

External actions on the system disturb the relationships between variables and thus should have a different impact on the other events. The application of conditioning is appropriate when an event occurs spontaneously (observation). It will lead to erroneous results if it is used to compute the effects of external actions: forcing the event to happen (intervention).

The "do" operator was introduced by Goldszmidt and Pearl (1992) for the ordinal conditional functions of Spohn (1988). Pearl (2000) introduced this tool in the Bayesian networks framework to compute the effect of interventions.

The action $do(a_i)$ in a causal model corresponds to a minimal perturbation of the existing system that forces the variable A_i to take the value a_i.

The idea behind the "do" operator, is to remove from the decomposition of the joint distribution the element that corresponds to the variable concerned by the intervention and replace each occurrence of this variable with its value. Therefore, the effect of an intervention $do(a_i)$ on the joint distribution is computed as follows:

$$P(a_1, \ldots, a_n | do(a_i)) = \frac{P(a_1, \ldots, a_n)}{P(a_i | Pa(A_i))} = P(a_1, \ldots, a_n | a_i, Pa(A_i)) \cdot P(Pa(A_i))$$

Hence, on a causal Bayesian network, we can model the effects of not only observations but also those of interventions.

Example 2.7. *Let us consider the causal Bayesian network shown in Figure 2.5. Let us compare between the effect of seeing that someone has yellow teeth (i.e., the variable T takes spontaneously the value t_1) and the effect upon acting on the variable T, by painting one's teeth yellow (forcing T to take the specific value t_1).*

- *The effect of observing T with the value t_1 is computed with:*
  $P(s_1, t_1 | see(t_1))$
  $= P(s_1, t_1 | t_1)$
  $= \dfrac{P(s_1, t_1)}{P(t_1)} = \dfrac{0.36}{0.48} = 0.75$
- *The effect of acting on T forcing it to take the value t_1 is computed with:*
  $P(s_1, t_1 | do(t_1))$
  $= \dfrac{P(s_1, t_1)}{P(t_1 | Pa(t_1))} = \dfrac{P(s_1, t_1)}{P(t_1 | s_1)}$
  $= P(s_1) = 0.4$

You notice that intervening on the variable T affects differently the system.

Graphical representation of interventions

To handle interventions, changes on the structure of the causal Bayesian network are made. Two methods were developed namely, the mutilation and the augmentation of the causal

graph. These approaches have been proved to be equivalent under the probabilistic framework (Pearl, 2000).

Graph mutilation Since an intervention on a variable should not change our beliefs on its direct causes (parents), all the edges pointing to the node concerned by the action will be deleted, no changes affect the causes of the target variable. The altered graph is called a mutilated graph denoted by $\mathcal{G}_{mut}$.

It is considered as a surgery (a mutilation) by which all the other causes than the one of the intervention will be excluded.

Let $\mathcal{G} = (V, E)$ be a causal network on which we make an intervention on a variable $A_i \in V$ by forcing it to take the value a_i ($\text{do}(a_i)$). We define mutilation on two steps:

1. Arcs pointing to A_i in $\mathcal{G}$ will be deleted. The obtained mutilated graph is denoted $\mathcal{G}_{mut}$. Its associated distribution is denoted $P_{\mathcal{G}_{mut}}$. This intervention affects the computation of the joint distribution $P_{\mathcal{G}}$ by transforming it into $P_{\mathcal{G}}(.|\text{do}(a_i))$. In the mutilated graph, it corresponds to observing $A_i = a_i$, i.e., applying the definition of conditioning.

$$P_{\mathcal{G}_{mut}}(.|a_i) = P_{\mathcal{G}}(.|do(a_i)) \tag{2.6}$$

2. An action $do(a_i)$ imposes a value a_i on a variable A_i. Accordingly, the corresponding distribution of A_i becomes:

$$P(a_k) = \begin{cases} 1 & \text{if } a_k = a_i \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

Example 2.8. *Let us continue with the causal network presented in Figure 2.5. By painting one's teeth yellow (forcing T to take the specific value t_1), the state of T will be independent from the fact of smoking (S). Therefore, the link relating S to T will be deleted. This is represented by the graph in Figure 2.6.*

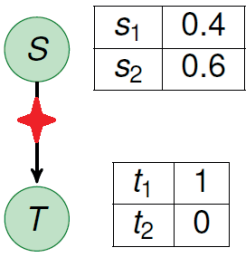


Figure 2.6: Graph mutilation upon the intervention $do(t_1)$

Graph augmentation Another interpretation of interventions on a causal network, is to add a fictive parent node called "DO" to the variable A_i on which an intervention is performed. The added node is considered as an extra node in the system. The set of the parents of A_i becomes $PA' = PA \cup DO$. The altered structure is called in augmented graph and denoted by $\mathcal{G}_{aug}$.

The DO node is taking value in $do(x)$, $x \in \{\Theta_{A_i} \cup \{\text{nothing}\}\}$. $do(nothing)$ means that there are no actions on the variable A_i, it represents the state of the system when no interventions are made. $\text{do}(a_i)$ means that the variable A_i is forced to take the value a_i. Hence, the graph augmentation method allows to represent the effect of observations and interventions.

The new distribution of the node DO is defined as:

$$P(a_k|Pa(A_i), do(x)) = \begin{cases} 1 & \text{if } x = a_i \\ 0 & \text{if } x \neq a_i \\ P(a_k|Pa(A_i)) & \text{if } x = nothing \end{cases} \tag{2.8}$$

It remains to specify what is the distribution assigned to the added node (i.e., DO). Two cases are considered:

- If there is no intervention then the distribution of the DO node is defined by:

$$P(do(x)) = \begin{cases} 1 & \text{if } x = nothing \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

- If there is an intervention that pushes the variable A_i to take the value a_i, then the distribution relative to DO is defined by:

$$P(do(x)) = \begin{cases} 1 & \text{if } x = a_i \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

Example 2.9. *Let us continue with the causal network presented in Figure 2.5. By painting one's teeth yellow, we are forcing the variable T to take the specific value t_1. The set of parents of T becomes $S \cup DO$ where the node DO is set to the value $do(t_1)$. This is represented by the graph in Figure 2.7.*

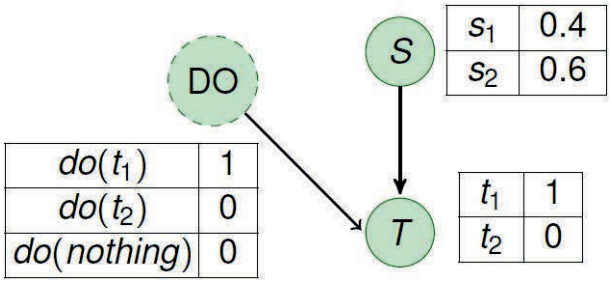


Figure 2.7: Graph augmentation upon the intervention $do(t_1)$

Semi Markovian Causal Models

Causal Bayesian networks (Pearl, 2000) are used to represent and reason with probabilistic causal knowledge. As we have seen the concept of intervention is very important in these graphical models for causal analysis.

To ascribe causality by integrating *latent* variables, causal Bayesian networks were extended to Semi Markovian Causal Models (Tian & Pearl, 2002). In this latest, causal reasoning is based on observed as well as unobserved variables.

Let us denote by $V = \{V_1, \ldots, V_n\}$ the set of observed variables and by $U = \{U_1, \ldots, U_n\}$ the set of unobserved variables. If no U variable is a descendant of any V variable, then the corresponding model is called a Semi Markovian model.

In a Semi Markovian model, the observed probability distribution, $P(v)$, becomes a mixture of products:

$$P(v) = \sum_{u} \prod_{i} P(v_i|pa_i, u^i) \cdot P(u)$$

where PA_i and U^i stand for the sets of the observed and unobserved parents of V_i, and the summation ranges over all the U variables.

Causal possibilistic network

Aside from causal Bayesian networks, causal possibilistic networks (Benferhat, 2010; Benferhat & Smaoui, 2011) were developed for cases where the causal knowledge is expressed under the possibilistic framework. They are especially appropriate tools to fit pure qualitative knowledge when only the ordinal handling is important.

A causal possibilistic network has the same structure as an associational possibilistic network whether defined on the qualitative or the quantitative setting (see subsection 1.6.3). However, as for causal Bayesian networks arcs between variables are not only representing dependencies but also direct causal relations.

To compute the effect of an intervention, the counterpart of the "do" operator to the possibilistic framework has been defined (Benferhat & Smaoui, 2007a). Besides, it was shown that the graph mutilation and the graph augmentation methods are equivalent approaches to graphically represent interventions under the possibilistic framework.

2.7 Causality as a nonmonotonic relation

2.7.1 Non monotonic relation

The relation of nonmonotonic inference to causality has already been emphasized by researchers dealing with reasoning about actions and the frame problem (e.g., (Giunchiglia et al., 2004)).

It is an alternative to the model based on structural equations (Halpern & Pearl, 2005), which may be impossible to construct in situations of poor information. This was motivated by the insufficiency of material implication for representing a causal link between a cause and an effect.

Nonmonotonic reasoning handles rules that may have exceptions and deal with incomplete information. Thus, conclusions are revised as further information becomes available.

2.7.2 Nonmonotonic consequence approach

The nonmonotonic consequence relation, "In context e_j, e_i is generally true" is denoted $e_j \mid\!\sim e_i$, where e_j and e_i are respectively instances of the variables E_j and E_i. It can be viewed as a constraint stating that $e_j \wedge e_i$ is strictly more plausible than $e_j \wedge \neg e_i$.

The nonmonotonic consequence relation $\mid\!\sim$, satisfies the requirements of System P (P stands for preferential) (Kraus et al., 1990), which is considered as a common kernel of several nonmonotonic formalisms, that consist in:

- *Right Weakening:* E $\mid\!\sim$ F and F $\models$ G imply E $\mid\!\sim$ G;
- *Left AND:* E$\mid\!\sim$F and E$\mid\!\sim$ G imply E$\mid\!\sim$F$\wedge$ G;
- *Right OR:* E$\mid\!\sim$G and F$\mid\!\sim$ G imply E$\vee$ F$\mid\!\sim$ G;
- *Cautious monotony:* E$\mid\!\sim$F and E$\mid\!\sim$G imply E$\wedge$ F$\mid\!\sim$G;
- *Cut:* E$\mid\!\sim$F and E$\wedge$F$\mid\!\sim$G imply E$\mid\!\sim$G.

In some cases, a strong version of Cautious Monotony is considered, namely the property of Rational Monotony (Lehmann & Magidor, 1992). It is defined as:

- *Rational Monotony:* E $\not\mid\!\sim$F and E$\mid\!\sim$G imply E$\wedge$ F$\mid\!\sim$ G

Nonmonotonic relations have also been represented in different uncertainty frameworks. In fact, some models are based on probability theory (Adams, 1975), possibility theory (Benferhat et al., 1992) as well as the theory of belief functions (Benferhat et al., 2000; Smets & Hsia, 1991).

In (Bonnefon et al., 2006, 2008), a framework for ascribing causality from a background knowledge using a nonmonotonic relation (satisfying System P) has been proposed. The inference engine based on System P is very cautious. This approach, unlike the one using structural equations (Halpern & Pearl, 2005) is based on qualitative information, which may be incomplete and tolerate exceptions. The aim of this approach is to predict causal relations that an agent may assign using available information and by distinguishing normal events from abnormal.

Let us denote by C a given context and by A and B two events. If an agent believes that " in context C, B is generally true", then the relation $C \mid\!\sim B$ holds. Ignorance, namely " in context C, B is ignored" is formalized with $C \not\mid\!\sim B$ and $C \not\mid\!\sim \neg B$.

Note that beliefs are only temporarily accepted since they can be questioned after a new piece of information become available. Thus, the ascription of causality, facilitation as well as the identification of justification links takes into account the dynamic of beliefs.

Causality ascription

Given the assertion "In context C, B is known to be exceptional" ($C \mid\!\sim \neg B$). If after that the event A takes place, B becomes generally true (C $\wedge$ A$\mid\!\sim$B), then in context C, A is perceived to be the cause of B.

Definition 2.2. *A is a cause of B in context C,* $A \Rightarrow_{ca} B$ *, if* $C \mid\sim \neg B$ *and* $C \wedge A \mid\sim B$.

Example 2.10. *Assume that in a given context C, having yellow teeth T is an exceptional event:* $C \mid\sim \neg T$. *If the event S (smoking) occurs, then T becomes a normal event,* $C \wedge S \mid\sim T$. *In this case, smoking is seen to be the cause of having yellow teeth, i.e.,* $C : S \Rightarrow_{ca} T$.

Facilitation ascription

Facilitation requires less beliefs from the agent that causality. Assume that in a given context C, the occurrence of the event B is known to be exceptional (and observed to be false, ($C \mid\sim \neg B$)). If the event A is reported along with B which becomes ignored (($C \wedge A \mid\sim \neg B$) and (C$\wedge A \mid\sim B$)), we say that in context C, A is perceived to have facilitated the occurrence of B since the occurrence of A makes the occurrence of B (i.e., B is true) unsurprising (but not expected) to the agent.

Definition 2.3. *A is a facilitation of B in context C,* $A \Rightarrow_{fa} B$ *, if* $C \mid\sim \neg B$ *and* $C \wedge A \not\mid\sim \neg B$ *and* $C \wedge A \not\mid\sim B$.

Example 2.11. *In context C, one has generally no Cervical intraepithelial neoplasia (CIN),* $C \mid\sim \neg CIN$. *This is no longer true when smoking (S) (*$C \wedge S \not\mid\sim \neg CIN$*), even though it does not systematically or almost systematically generate CIN (*$C \wedge S \not\mid\sim CIN$*). Suppose now that a person got CIN, this person is a smoker. Smoking will be judged as having facilitated CIN, i.e.,* $C : S \Rightarrow_{fa} CIN$.

Justification ascription

Justification is related to the notion of explanation following Spohn. If in context C, the state of B is ignored ($C \not\mid\sim B$) and ($C \not\mid\sim \neg B$) and after the occurrence of the event A it becomes true ($C \wedge A \mid\sim B$), A gives reason to expect or to justify the occurrence of B.

Definition 2.4. *A is a justification of B in context C,* $A \Rightarrow_{ju} B$*, if* $C \not\mid\sim B$ *and* $C \not\mid\sim \neg B$ *and* $C \wedge A \mid\sim B$.

Example 2.12. *Giving up smoking is an ignored event. Formally, we have* $C \not\mid\sim \neg S$ *and* $C \not\mid\sim S$. *If after knowing that the smoker took nicotine patches (P), the event giving up smoking becomes true (*$C \wedge P \mid\sim \neg S$*). Taking nicotine patches will be seen as justifying stopping smoking, i.e.,* $C : P \Rightarrow_{ju} \neg S$.

2.7.3 Trajectory-based preference relations

This approach starts with the idea of counterfactual causes. It involves two computation concerning the evolution of the world as follows:

- Extrapolation (Dupin de Saint-Cyr & Lang, 2002): it consists in computing the most normal evolutions of the world (called trajectories). It is a process of completing initial beliefs from observations by assuming minimal abnormalities in the evolution of the world with respect to generic knowledge.
- Update (Katsuno & Mendelzon, 1991): it aims at identifying a minimal change with respect to the initial scenario. The update operator proposed in (Dupin de Saint-Cyr, 2008) is based on a distance between trajectories that take into account the time point of the change and normality.

To discriminate between potential counterfactual causes, we will choose among the normal ones, the most abnormal ones in context.

Example 2.13. *The event $Smoking_t$ that took place at time t caused* Lung *$cancer_{t+N}$ that occurred at time $t+N$ in a given scenario if:*

- *$Smoking_t$ and* Lung *$cancer_{t+N}$ took place in this scenario.*
- *if $Smoking_t$ had not taken place in this scenario, then* Lung *$cancer_{t+N}$ would not have taken place (counterfactuality).*

For that we have to:

- *compute if $Smoking_t$ and* Lung *$cancer_{t+N}$ take place in scenario (extrapolation)*
- *compute what would have happened if $Smoking_t$ had not taken (updating by $\neg Smoking_t$)*

2.7.4 Norm-based approach

This approach (Kayser & Mokhtari, 1998; Khelfallah & Mokhtari, 2001; Kayser & Nouioua, 2005) is based on norms to define causality where norms are rules that apply by default and the knowledge necessary to causal ascription is expressed in a reified first-order logic augmented with default rules.

In this causal approach when the event is considered normal, its cause is the norm itself. However, if it is abnormal, its cause is traced back to the violation of a norm.

For that looking for the cause of an abnormal event E occurring at time t basically amounts to finding an agent who should, according to some norm, adopt behavior b at a time $t' < t$, and actually adopted another behavior b', such that E appears as a normal consequence of b'.

It also requires that, at t', the agent had the possibility to have the normal behavior b; otherwise, b' is only a derived anomaly and the search must be pursued to find a primary anomaly, occurring earlier than t' and explaining the impossibility of the agent to have the behavior b at t'. In the case where this search fails, non agentive abnormal circumstance that could explain E should be identified.

The fact that property P holds for agent A at time t is written: $holds(P, A, t)$. Two modalities are introduced to express norm violations: $should(P, A, t)$ and $able(P, A, t)$ standing for: at time t, A should (resp. has the ability to) achieve P.

However, the generalization of this approach to domains where norms are only what is normal norms should be organized in a hierarchy. Besides, the most violated norm should be perceived as the cause of an abnormal event. The issue is that verifying this hypothesis requires to gather a reasonably complete set of norms for the domain, which is a hard task (Benferhat et al., 2008).

Example 2.14. *This example is extracted from (Benferhat et al., 2008)). Let us consider these literals:*

- *Wet* $\rightarrow$ *should(reduced_speed,A, t)*
- *holds(Acc,A,t)* $\rightarrow$ *should(avoid_obs, A,*$t-1$*)*
- *should(avoid_obs,A,t)*$\rightarrow \neg$ *able(ch_lane,A,t)* $\wedge$ *should(stop,A,t)*

Expressed in this language, the cause of an abnormal event (the primary anomaly P_ano) obtains as:

- *should(F,A,t)* $\wedge$ *able(F,A,t)*$\rightarrow \neg$ *holds(F,A,*$t+1$*)* $\wedge$ *P_ano(F,A,*$t+1$*)*

It means that if at t *an agent* A *should do* F *and was able to do* F*, while at* $t+1$*,* F *failed to be done, this failure is the cause looked for.*

2.8 Conclusion

In this chapter, we gave an overview of the different definitions of causality. Under uncertainty, we presented a brief review of the approaches. We explained how to ascribe causality under a probabilistic framework. Besides, we clarified the relationship between associational networks and causal networks under the probabilistic and possibilistic framework. Indeed, these latter allow to handle interventions via the "do" operator. Besides, we explained that the nonmonotonic approaches allows a consistent and efficient ascription of causality.

In order to introduce the causal belief network that will be used to formalize the imperfect causal knowledge, we will first in the next chapter analyze and revise the theoretical foundations of existing belief networks with conditional beliefs and propose a new associational graphical model under the belief function framework.

Part II

Modeling causality under the belief function framework

Chapter 3

A new associational belief network

3.1 Introduction

The belief function theory is a powerful tool to handle imperfect knowledge. The representation and reasoning from a belief function knowledge is simply done on graphical models. Through direct acyclic graphs (DAGs), it is possible to compactly encode a collection of (in)dependence statements and to predict the effect of observations on the joint distribution. As for Bayesian networks, the structure of these associational networks will serve as a basis to model causality.

Existing associational belief networks (known also as evidential networks) are confusing. This is due to the misleading sense of the term evidential networks. In fact, this word is originally coined for valuation networks and is nowadays used for both directed acyclic graphs and valuation networks. Indeed, no consensus exists about a unique definition and representation of belief networks as DAGs.

As mentioned in Chapter 1, belief networks can be categorized into two main groups as follows: those where conditional dependencies can be defined by a unique conditional distribution for each child node given all its parents like for Bayesian networks (Simon et al., 2008) and those where conditional distributions relating a node to each of its parents are specified (Ben Yaghlane & Mellouli, 2008; Xu & Smets, 1996). In this latter, conditional distributions can be aggregated into one distribution representing the relation between a node and all its parents.

To compute the global joint distribution, beliefs defined for each node in the context of its parents should be aggregated. The conjunctive rule of combination in its basic form (as well as Dempster's rule of combination for normalized beliefs) is used to fusion non-conditional beliefs defined on the same frame of discernment. Accordingly, to use these rules the ballooning extension is first applied to get rid of conditioning then deconditionalized beliefs are vacuously extended. However, using these rules will not truly reflect the initial knowledge.

In the first part of the this chapter, we propose an analysis of existing belief networks modeled with DAGs. Besides, since we have emphasized that the belief function theory is a

generalization of probability theory when focal elements are singletons, we look in this chapter if a belief network collapses into a Bayesian network in the case of Bayesian *bba*s. In the second part, we present a graphical model based on new ways to get rid of conditioning and to extend beliefs to a product space.

The rest of this chapter is organized as follows: in Section 3.2, we analyze the theoretical foundations of existing belief networks. Section 3.3 investigates the relationship between belief networks and Bayesian networks. Section 3.4 presents our belief network where beliefs are expressed in terms of conditional mass distributions.

3.2 Analysis of existing belief networks

This section reviews existing associational networks with conditional beliefs where knowledge is formalized with the belief function theory. We will investigate the used operators and check if they lead to a loss of information. Then, we will compare between the different used chain rules and examine if they lead to the same global joint distribution. Examples will be given to illustrate the weaknesses of these models.

3.2.1 Belief networks: conditionals for all parents

Evidential networks (EN)

The recent so-called *Evidential Network (EN)* (Simon et al., 2008) is a belief network that is similar to a Bayesian network. It combines the belief function theory with a directed acyclic graph (DAG). The idea is to encode a mass as a probability on a powerset of domains associated with each variable. More precisely, an evidential network is defined on two levels:

- *Qualitative level:* represented by a DAG, $G = (V, E)$ where V is the set of variables, V and E is the set of edges encoding the dependencies among variables. A given variable A_i is called a parent of variable A_j if there is an arc from A_i to A_j. The set of parents of A_i is denoted by $PA(A_i)$. A subset from $PA(A_i)$ is denoted by $Pa(A_i)$.
- *Quantitative level:* represented by the set of *bba*s associated to each node in the graph.
  - For each root node A_i (i.e., node without parent nodes) having a frame of discernment Θ_{A_i}, an a priori mass distribution m^{A_i} is defined on the powerset $2^{\Theta_{A_i}}$.
  - For other nodes, a conditional *bba* $m^{A_i}(.|Pa(A_i))$ is specified for each value of A_i knowing the value of all the parents $Pa(A_i)$. It means that a conditional belief mass table represents the relation between the basic belief masses expressed on the frame of discernment of the child node and the basic belief masses defined on the frame of discernment of the variables in the parent nodes.

This kind of network (Simon et al., 2008) is proposed in context of *reliability modeling* where the domain associated with variables is defined as {Up, Down}.

The proposed network can be seen as a standard Bayesian network where each node can take one of these three values {Up}, {Down}, {Up, Down}. The possibility to allocate a quantity to the proposal {Up, Down} softens Bayesian networks.

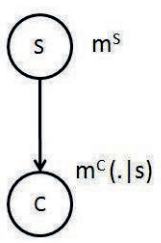
The subset {Up, Down} models the case of total ignorance on the real state of the component without commitment. It means that the component can be in the state {Up} or {Down}. However, the way the disjunction {Up, Down} is used reflects that this value is rather considered as a new value of the frame of discernment. Accordingly, the EN is seen a Bayesian network where the probabilistic frame of discernment has been changed from $\{\theta_1, \theta_2\}$ to $\{\theta_1, \theta_2, \theta_3\}$ where $\theta_1 = \{Up\}$, $\theta_2 = \{Down\}$ and $\theta_3 = \{Up, Down\}$.

Global joint distribution computation

As for Bayesian networks, the product of conditionals *bba*s for each variable provides a unique joint mass distribution for all variables in the network that agrees with the independencies represented by the network structure. Hence, the global joint mass distribution is computed by multiplying all conditional masses as follows:

$$m^V = \prod_{i=1}^{n} m^{A_i}(.|Pa(A_i)) \tag{3.1}$$

Example 3.1. *Let us consider the belief network presented in Figure 3.1.*



$\{s_1\}$	0.1
$\{s_2\}$	0.7
Θ_S	0.2

	$\{s_1\}$	$\{s_2\}$
$\{c_1\}$	0.8	0.1
$\{c_2\}$	0.1	0.7
Θ_C	0.1	0.2

Figure 3.1: Evidential network (EN)

Nodes are described as follows:

- *S describes the presence of sugar in the cup of coffee, $\Theta_S = \{s_1, s_2\}$ where s_1 is yes and s_2 is no.*
- *C represents the sweetness of the coffee, $\Theta_C = \{c_1, c_2\}$ where c_1 is sweet and c_2 is bitter.*

The joint distribution is computed by making the cartesian product of the conditional masses as follows:

$$\forall s \subseteq \Theta_S, c \subseteq \Theta_C, m(s \times c) = m(s) \cdot m(c)$$

Note that total ignorance about the state of C is represented with the vacuous bba $m^S(.|\Theta_C) = 1$.

Results after the computation of the global joint distribution are presented in Table 3.1.

Table 3.1: Global joint distribution in EN: m^{SC}

$\{(s_1, c_1)\}$	0.08
$\{(s_1, c_2)\}$	0.01
$\{s_1\} \times \Theta_C$	0.01
$\{(s_2, c_1)\}$	0.07
$\{(s_2, c_2)\}$	0.49
$\{s_2\} \times \Theta_C$	0.14
$\Theta_S \times \{c_1\}$	0
$\Theta_S \times \{c_2\}$	0
$\Theta_S \times \Theta_C$	0.2

According to this chain rule (Equation 3.1), conditioning is similar to the one defined for probability theory where for $a_1 \subseteq \Theta_{A_1}$ and $a_2 \subseteq \Theta_{A_2}$ we have:

$$m^{A_1}(a_1|a_2) = \frac{m^{A_1,A_2}(a_1 \cap a_2)}{m^{A_2}(a_2)} \tag{3.2}$$

However, as proved in (Vejnarová, 2012), it is not possible to consider such conditioning with basic assignments. It is caused by a simple fact that the mass function m, in contrary to *bel* and *pl* is not monotonous with respect to set inclusion.

Counter-example 3.1. *Let us consider the frame* $\Theta_{A_1} = \{a_{11}, a_{12}, a_{13}\}$ *where basic belief masses are defined as* $m^{A_1}(a_{11}) = 0.7$, $m^{A_1}(\Theta_{A_1}) = 0.3$.

To compute $m^{A_1}(a_{11}|\Theta_{A_1})$ *using Equation 3.2 will give:*

$\frac{m^{A_1}(a_{11})}{m^{A_1}(\Theta_{A_1})} = \frac{0.7}{0.3} = 2.33$. *However, a bbm cannot exceed one.*

For inference, the evidential network (EN) uses the junction tree algorithm and the total probability theorem. Clearly, there is no difference between this network and Bayesian networks in terms of representation and also propagation of beliefs. However, as proved in (Laâmari et al., 2010), the use of the junction tree (JT) algorithm is computationally less efficient than the use of tools specific to the belief function theory (e.g., the modified binary joint tree (MBJT) (Ben Yaghlane & Mellouli, 2008)). Indeed, using probabilistic tools may present inconveniences. In fact, the used inference algorithm satisfies the additive axiom inherent to Bayesian network algorithm which does not exist in the general framework of the belief function. As a consequence, this network cannot serve as a basis for the causal belief network.

3.2.2 Belief networks: conditionals per single parent

Unlike Bayesian networks, conditional distributions are defined per edge in these belief function networks. Two representations with conditional belief distributions exist and will be analyzed here namely the *Evidential Networks with Conditional beliefs (ENC)* and the *Directed EVidential Networks (DEVN).*

Evidential networks with conditional beliefs (ENC)

The so-called *Evidential Networks with Conditional beliefs* (Smets, 1993b; Xu & Smets, 1996) have *binary* relations between nodes of the graphical model. Besides, in these networks

conditional belief distributions are defined in a different way from the one traditionally used in Bayesian networks and in Evidential networks (EN). Actually, an ENC is defined on two levels:

- *Qualitative level:* represented by a directed acyclic graph (DAG), $G = (V, E)$. If there is an edge from variable A_i to variable A_j, A_i is called a parent of A_j.
- *Quantitative level:* represented by the set of belief functions (*bel*) associated to each *edge* in the graph. An edge represents a conditional relation between the two nodes it connects. The knowledge about the relations between two nodes can be issued from different sources (local conditional beliefs).

The number of conditional relations in the ENC, depends on the number of edges whereas in EN it depends on the number of child nodes. The main advantage of defining conditionals per single parent is to simplify knowledge acquisition and storage. Figure 3.2 shows a network with individual parents.

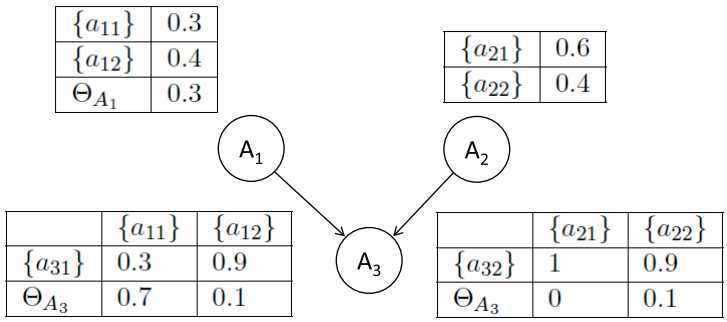


Figure 3.2: A belief network with individual parents

However in the ENC, if a conditional belief is given in the context of more than one parent node, then these nodes should be merged into one node.

Example 3.2. *Let us consider a DAG with three nodes A_1, A_2 and A_3 where A_1 and A_2 are parents of A_3. Imagine that experts, give the following conditional distribution $m^{A_3}(.|a_1, a_2)$ where $a_1 \subseteq \Theta_{A_1}$ and $a_2 \subseteq \Theta_{A_2}$. In this case, the node A_1 and A_2 should be* merged into one node *(see Figure 3.3).*

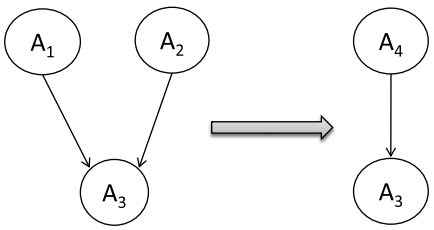


Figure 3.3: Evidential network with conditional belief

This associational network is not adequate to serve as a basis to model causality. In fact, despite its storage power, it lacks from flexibility to model knowledge. Besides, it leads to a loss of information since a priori beliefs about parent nodes will be lost after the fusion process.

Directed evidential networks (DEVN)

To tackle the issue of binary relations encountered with *ENC*s, *Directed EVidential Networks (DEVN)* (Ben Yaghlane & Mellouli, 2008) were proposed. In these latter, relations have been generalized to many nodes. Accordingly, no need to merge parent nodes in one single node if beliefs are expressed given all parents. Besides, in *DEVN*s conditional belief distributions can be defined either per single parent or for all parents. Obviously, they offer more flexibility. It is defined in two levels:

- Qualitative level: a directed acyclic graph (DAG) in which the nodes represent variables, and directed arcs describe the conditional dependence relations embedded in the model.
- Quantitative level: dependence relations are expressed by conditional belief functions (*bel*) for each node in the context of either one parent node or for all parents.

Global joint distribution computation

The definition of the global joint belief distribution is different from the one used for Bayesian networks and Evidential networks (*EN*). In fact, all local knowledge are aggregated using the conjunctive rule of combination (or Dempster's rule of combination).

The chain rule is computed either for *ENC* or *DEVN* with belief function distributions (*bel*) as follows:

$$bel^{A_1,\ldots,A_n} = \bigcirc\!\!\!\!\!\!\cap_{i=1,\ldots,n} \; \bigcirc\!\!\!\!\!\!\cap_{\omega \in Pa(A_i)} bel^{A_i}(.|\omega)^{\uparrow A_i \times Pa(A_i)} \tag{3.3}$$

Unfortunately, this formula is not clear. Indeed, the following points present some issues:

- the ballooning extension was defined for basic assignments (Shafer, 1982) (and for plausibility functions (Foucher et al., 2006)) and it is not clear how it will be applied directly on belief functions unless beliefs will to be transformed to mass functions using möbius transformation.
- since *DEVN* allows to define conditional belief distributions per edge or for all parents, it is not clear within the following rule, if ω is an instance from a single parent or it represents a node from the set of the parents of a node A_i.
- it is not clear if the index i concerns the number of parents or the cardinality of the variable of interest A_i.
- it is not obvious how we can apply the conjunctive rule of combination on beliefs defined on disjoint frames.

Please note that experts usually express their beliefs with mass distributions. Thus, we propose to clarify the proposed chain rule when beliefs are defined with basic assignments (Boukhris, Elouedi, & Benferhat, 2011a).

To compute the global joint belief distribution, the ballooning extensions for the deconditionalization process are first computed for local conditional distributions. All deconditionalized beliefs are vacuously extended and then aggregated. To better explain this process, we will use these following notations:

- $PA(A_i)$: the set of the parents of A_i;
- $Pa(A_i)$: a subset from $PA(A_i)$;
- $PA_j(A_i)$: a single parent of A_i;
- $Pa_j(A_i)$: a subset from $PA_j(A_i)$.

In the general case, i.e., where beliefs are given per single parent, the computation of the global joint distribution is done in three steps:

1. For a conditional variable A_i:

   1.1 For each subset of a single parent denoted by $Pa_j(A_i)$, compute the ballooning extension of $m^{A_i}(.|Pa_j(A_i))$ for the deconditionalization process:

   $$m^{A_i}(.|Pa_j(A_i))^{\uparrow A_i \times PA_j(A_i)}$$

   1.2 Combine the deconditionalized beliefs using the conjunctive rule of combination.

   $$\bigcirc\!\!\!\!\!\cap_{Pa_j(A_i)} m^{A_i}(.|Pa_j(A_i))^{\uparrow A_i \times PA_j(A_i)}$$

2. Extend each node (root node and child node) to the universe of all the variables in the network by applying the vacuous extension.

   $$(\bigcirc\!\!\!\!\!\cap_{Pa_j(A_i)} m^{A_i}(.|Pa_j(A_i))^{\uparrow A_i \times PA_j(A_i)})^{\uparrow A_1 \times \ldots \times A_n}$$

3. Combine local joint distributions using the conjunctive rule of combination and thus get the following chain rule:

$$m^{A_1,\ldots,A_n} = \bigcirc\!\!\!\!\!\cap_{i=1,\ldots,n} (\bigcirc\!\!\!\!\!\cap_{Pa_j(A_i)} m^{A_i}(.|Pa_j(A_i))^{\uparrow A_i \times PA_j(A_i)})^{\uparrow A_1 \times \ldots \times A_n} \tag{3.4}$$

As you may notice, this belief chain rule is different from the one based on the product operator that is proposed for evidential networks (Simon et al., 2008).

Example 3.3. *Let us continue with the belief network presented in Figure 3.1. Its corresponding global joint distribution is computed using the three steps exposed above. Namely,*

1. *Ballooning extension of each conditional mass:*

   - *Deconditionalization for each instance of the parents:*
     To get rid from conditioning, we have to compute the ballooning extension (Equation 1.36) of each instance of the parents ($m^C(.|s_1)$ and $m^C(.|s_2)$).
     For example, the bbm $m^C(c_1|s_1)$ will be transferred to $\{(c_1, s_1)\} \cup \{\Theta_C \times \overline{s_1}\}$, i.e., to $\{(s_1, c_1), (s_2, c_1), (s_2, c_2)\}$. Results are presented in Table 3.2.

Table 3.2: Ballooning extensions of $m^C(.|s_i)$

$m^C(.\|s_1)^{\uparrow SC}$	$\{(s_1,c_1),(s_2,c_1),(s_2,c_2)\}$	0.8
	$\{(s_1,c_2),(s_2,c_1),(s_2,c_2)\}$	0.1
	$\Theta_S \times \Theta_C$	0.1
$m^C(.\|s_2)^{\uparrow SC}$	$\{(s_2,c_1),(s_1,c_1),(s_1,c_2)\}$	0.1
	$\{(s_2,c_2),(s_1,c_1),(s_1,c_2)\}$	0.7
	$\Theta_S \times \Theta_C$	0.2

- *Aggregate deconditionalized beliefs using the conjunctive rule of combination:*
  The deconditionalized bbms are defined on the same frame of discernment namely $\Theta_S \times \Theta_C$. Their combination using the conjunctive rule of combination gives the local joint distribution at the node C.
  To combine for instance $m^C(c_1|s_1)$ with $m^C(c_1|s_2)$, we have to identify the intersection of the subsets given by their respective ballooning extension:
  $\{(s_1,c_1),(s_2,c_1),(s_2,c_2)\} \cap \{(s_2,c_1),(s_1,c_1),(s_1,c_2)\}$
  $= \{(s_1,c_1),(s_2,c_1)\}$.
  Their product-intersection is computed afterwards as follows:
  $m_{s_1}^{C\uparrow SC}(\{(s_1,c_1),(s_2,c_1),(s_2,c_2)\}) \cdot m_{s_2}^{C\uparrow SC}(\{(s_2,c_1),(s_1,c_1),(s_1,c_2)\})$
  $= m^{SC}(\{(s_1,c_1),(s_2,c_1)\})$
  $= 0.8 \cdot 0.1 = 0.08$.
  Results are shown in Table 3.3.

Table 3.3: Local joint distribution at the node C

$m_{s_2}^{C\uparrow SC}$ \ $m_{s_1}^{C\uparrow SC}$	$\{(s_1,c_1),(s_2,c_1),(s_2,c_2)\} = 0.8$	$\{(s_1,c_2),(s_2,c_1),(s_2,c_2)\}$=0.1	$\Theta_S \times \Theta_C$ =0.1
$\{(s_2,c_1),(s_1,c_1),(s_1,c_2)\}$ =0.1	$\{(s_1,c_1),(s_2,c_1)\}$ = 0.08	$\{(s_1,c_2),(s_2,c_1)\}$ = 0.01	$\{(s_1,c_1),(s_2,c_1),(s_1,c_2)\}$ = 0.01
$\{(s_2,c_2),(s_1,c_1),(s_1,c_2)\}$ =0.7	$\{(s_1,c_1),(s_2,c_2)\}$ = 0.56	$\{(s_1,c_2),(s_2,c_2)\}$ = 0.07	$\{(s_1,c_1),(s_2,c_2),(s_1,c_2)\}$ = 0.07
$\Theta_S \times \Theta_C$ =0.2	$\{(s_1,c_1),(s_2,c_2),(s_2,c_1)\}$ = 0.16	$\{(s_1,c_2),(s_2,c_2),(s_2,c_1)\}$ =0.02	$\Theta_S \times \Theta_C$ = 0.02

2. *Vacuous extension to a joint space:*

To make the combination of all local joint distributions and find their possible intersections, variables will be extended to the product space of all the variables. In this example, masses defined on Θ_S should be vacuously extended to $\Theta_S \times \Theta_C$ as shown in Table 3.4.

Table 3.4: Vacuous extension: $m^{S\uparrow SC}$

$\{s_1\} \times \Theta_C$	0.1
$\{s_2\} \times \Theta_C$	0.7
$\Theta_S \times \Theta_C$	0.2

3. Combination of all local joint distributions:

The aggregation of the vacuous extension of subsets of Θ_S (Table 3.4) with the deconditionalized bbas of node C (Table 3.3) determines the global joint distribution of the network. The intersection of subsets specifies elements of the global joint distribution (see Table 3.5). The induced global joint bba is obtained by making the sum of bbms relative to each focal element found in Table 3.5. Results are summarized in Table 3.6.

Table 3.5: Elements of the global joint distribution in *DEVN*

$m^{C\uparrow SC}$ \ $m^{S\uparrow SC}$	$\{(s_1,c_1),(s_1,c_2)\}$=0.1	$\{(s_2,c_1),\{s_2,c_2\}\}$=0.7	$\Theta_S \times \Theta_C$=0.2
$\{(s_1,c_1),(s_2,c_1)\}$=0.08	$\{(s_1,c_1)\}$ =0.008	$\{(s_2,c_1)\}$ =0.056	$\{(s_1,c_1),(s_2,c_1)\}$ =0.016
$\{(s_1,c_2),(s_2,c_1)\}$=0.01	$\{(s_1,c_2)\}$ =0.001	$\{(s_2,c_1)\}$ =0.007	$\{(s_1,c_2),(s_2,c_1)\}$ =0.002
$\{(s_1,c_1),(s_2,c_1),(s_1,c_2)\}$ =0.01	$\{s_1\} \times \Theta_C$ =0.001	$\{(s_2,c_1)\}$ =0.007	$\{(s_1,c_1),(s_2,c_1),(s_1,c_2)\}$ =0.002
$\{(s_1,c_1),(s_2,c_2)\}$=0.56	$\{(s_1,c_1)\}$ =0.056	$\{(s_2,c_2)\}$ =0.392	$\{(s_1,c_1),(s_2,c_2)\}$=0.112
$\{(s_1,c_2),(s_2,c_2)\}$=0.07	$\{(s_1,c_2)\}$ =0.007	$\{(s_2,c_2)\}$ =0.049	$\{(s_1,c_2),(s_2,c_2)\}$=0.014
$\{(s_1,c_1),(s_2,c_2),(s_1,c_2)\}$=0.07	$\{s_1\} \times \Theta_C$ =0.007	$\{(s_2,c_2)\}$ =0.049	$\{(s_1,c_1),(s_2,c_2),(s_1,c_2)\}$=0.014
$\{(s_1,c_1),(s_2,c_2),(s_2,c_1)\}$=0.16	$\{(s_1,c_1)\}$ =0.016	$\{s_2\} \times \Theta_C$ =0.112	$\{(s_1,c_1),(s_2,c_2),(s_2,c_1)\}$=0.032
$\{(s_1,c_2),(s_2,c_2),(s_2,c_1)\}$=0.02	$\{(s_1,c_2)\}$ =0.002	$\{s_2\} \times \Theta_C$ =0.014	$\{(s_1,c_2),(s_2,c_2),(s_2,c_1)\}$=0.004
$\Theta_S \times \Theta_C$=0.02	$\{s_1\} \times \Theta_C$ =0.002	$\{s_2\} \times \Theta_C$=0.014	$\Theta_S \times \Theta_C$ =0.004

Table 3.6: Global joint distribution in *DEVN*: m^{SC}

$\{(s_1,c_1)\}$	0.08
$\{(s_1,c_2)\}$	0.01
$\{s_1\} \times \Theta_C$	0.01
$\{(s_2,c_1)\}$	0.07
$\{(s_2,c_2)\}$	0.49
$\{s_2\} \times \Theta_C$	0.14
$\{(s_1,c_1),(s_2,c_1)\}$	0.016
$\{(s_1,c_2),(s_2,c_1)\}$	0.002
$\{(s_1,c_1),(s_2,c_1),(s_1,c_2)\}$	0.002
$\{(s_1,c_1),(s_2,c_2)\}$	0.112
$\{(s_1,c_2),(s_2,c_2)\}$	0.014
$\{(s_1,c_1),(s_2,c_2),(s_1,c_2)\}$	0.014
$\{(s_1,c_1),(s_2,c_2),(s_2,c_1)\}$	0.032
$\{(s_1,c_2),(s_2,c_2),(s_2,c_1)\}$	0.004
$\Theta_S \times \Theta_C$	0.004

Aside from that the belief chain rule proposed for EN is not the same as that proposed for ENC and for DEVN*, the global joint distribution computed on EN presented in Table 3.1 is different and has less elements from the one found on ENC or* DEVN *presented in Table 3.6. In fact, some subsets were not considered by applying the product operator.*

3.3 Relationship with Bayesian Networks

The belief function theory is well known as a generalization of probability and quantitative possibility theories. In this section, we investigate the relationships between belief networks when beliefs are expressed through Bayesian *bba*s and Bayesian networks.

3.3.1 Evidential networks: *EN*

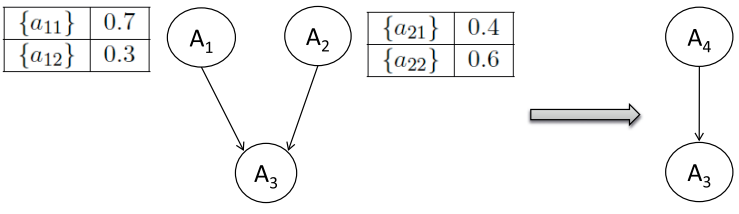
It is obvious that *EN*s are similar to Bayesian networks. Indeed, conditionals are defined for all parents and the product operator is used to aggregate knowledge. Unfortunately even if these networks show their efficiency in reliability applications, they cannot be applied in a general context since they are based on assumptions that are out of the framework of the belief function theory.

3.3.2 Evidential networks with conditional beliefs: *ENC*

Remember that in *ENC* (Xu & Smets, 1996) conditional beliefs are specified for each edge in the network. If beliefs are defined given all the parents, parents' nodes should be merged into one single node such that:

$$m^{A_i \times A_j}(C) = \begin{cases} m^{A_i}(A) \times m^{A_j}(B) & \text{if } C = A \times B, C \subseteq \Theta_{A_i} \times \Theta_{A_j}, A \subseteq \Theta_{A_i}, B \subseteq \Theta_{A_j} \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

Example 3.4. *Let us consider the belief function network shown in Figure 3.4. For seek of simplicity, we will assume that variables are binary. Let us suppose that beliefs in this network are defined for all parents (here A_3 given A_1 and A_2) with Bayesian bbas. According to the definition of an evidential network with conditional beliefs, relations between nodes should be binary. Therefore, nodes A_1 and A_2 should be merged into one single node denoted here by A_4.*



	$\{(a_{11}, a_{21})\}$	$\{(a_{11}, a_{22})\}$	$\{(a_{12}, a_{21})\}$	$\{(a_{12}, a_{22})\}$
$\{a_{31}\}$	0.7	0.4	0.2	0.9
$\{a_{32}\}$	0.3	0.6	0.8	0.1

Figure 3.4: Transformation of a network with Bayesian *bba* to an *ENC*

To compute the bba on the merged node A_4, beliefs of A_1 and A_2 should be aggregated using Equation 3.5. Since A_1 and A_2 are independent, elements of their cartesian product define the elements of A_4. Accordingly, to compute the a priori distribution $m^{A_1 \times A_2}$ (i.e., m^{A_4}), a pointwise product operation is made:

$$m^{A_1 \times A_2}(\{(a_{1i}, a_{2j})\}) = m^{A_1}(\{a_{1i}\}) \cdot m^{A_2}(\{a_{2j}\})$$

The mass distribution m^{A_4} is defined on $\Theta_{A_4} = \{a_{41}, a_{42}, a_{43}, a_{44}\}$ where $\{a_{41}\} = \{(a_{11}, a_{21})\}$, $\{a_{42}\} = \{(a_{11}, a_{22})\}$, $\{a_{43}\} = \{(a_{12}, a_{21})\}$, $\{a_{44}\} = \{(a_{12}, a_{22})\}$ as shown in Table 3.7:

Table 3.7: A priori *bba* after merging parents: m^{A_4}

$\{a_{41}\}$	0.28
$\{a_{42}\}$	0.42
$\{a_{43}\}$	0.12
$\{a_{44}\}$	0.18

The conditional distribution of the variable A_3 represented in Figure 3.4 is transformed to a conditional distribution in context of A_4 as shown in Table 3.8:

Table 3.8: Conditional *bba* after merging parents: $m^{A_3}(.|a_{4i})$

	$\{a_{41}\}$	$\{a_{42}\}$	$\{a_{43}\}$	$\{a_{44}\}$
$\{a_{31}\}$	0.7	0.4	0.2	0.9
$\{a_{32}\}$	0.3	0.6	0.8	0.1

Table 3.9 presents the global joint distribution after following the three steps presented in Subsection 3.2.2 leading to Equation 3.4.

These results are the same as those found by applying the product operator when considering this network as a Bayesian network. Note that when consonant bbas are defined, we found also the same results as those of a quantitative possibilistic network.

Table 3.9: Global joint distribution in ENC: m^{A_3, A_4}

$\{(a_{31}, a_{41})\}$	0.196
$\{(a_{31}, a_{42})\}$	0.168
$\{(a_{31}, a_{43})\}$	0.024
$\{(a_{31}, a_{44})\}$	0.162
$\{(a_{32}, a_{41})\}$	0.084
$\{(a_{32}, a_{42})\}$	0.252
$\{(a_{32}, a_{43})\}$	0.096
$\{(a_{32}, a_{44})\}$	0.018

Though the way to define conditional distributions is different from Bayesian networks, we have found the same results as those found on Bayesian networks. Defining conditionals per edge allows to reduce the storage memory space. However, the initial knowledge cannot be reconstructed after the fusion process.

On the other hand, a Bayesian network is equivalent to an *ENC* if for each child node, its parents are merged into one node. The a priori belief in this new node is the combination of the a priori of each parent node.

Accordingly, *ENC*s cannot be seen as true extensions of Bayesian networks. A counterexample is provided with their generalized form namely *DEVN*s.

3.3.3 Directed evidential networks with conditional beliefs: *DEVN*

As mentioned before, *DEVN* (Ben Yaghlane & Mellouli, 2008) allows to specify conditional beliefs for each edge in the network. However, by defining each conditional separately, we allow that the initial belief of the child node is *not unique*.

In fact, it depends on the chosen path which is not the case for Bayesian networks. For these reasons, we provide a counterexample where we show that *DEVN*s cannot be regarded as a true extension of Bayesian networks.

Counter-example 3.2. *Let us consider the network depicted in Figure 3.5 where basic assignments are provided per single parent.*

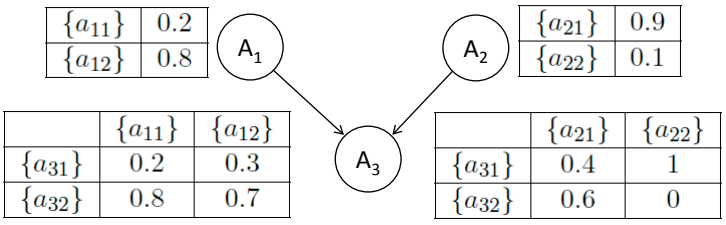


Figure 3.5: *DEVN* with Bayesian *bba*

This belief network is a DAG where all focal elements are singletons. However, there is no probability distribution that satisfies this graph when there is a mass distribution that does.

Results for the joint mass distribution on $\Theta_{A_1} \times \Theta_{A_3}$, $\Theta_{A_2} \times \Theta_{A_3}$ *are shown in Table 3.10.*

Table 3.10: Local joint mass distributions

$m^{A_1A_3}(\{(a_{11}, a_{31})\})$	0.04
$m^{A_1A_3}(\{(a_{11}, a_{32})\})$	0.16
$m^{A_1A_3}(\{(a_{12}, a_{31})\})$	0.24
$m^{A_1A_3}(\{(a_{12}, a_{32})\})$	0.56

$m^{A_2A_3}(\{(a_{21}, a_{31})\})$	0.36
$m^{A_2A_3}(\{(a_{21}, a_{32})\})$	0.54
$m^{A_2A_3}(\{(a_{22}, a_{31})\})$	0.1
$m^{A_2A_3}(\{(a_{22}, a_{32})\})$	0

After a projection on Θ_{A_3}, *we notice that the distributions of* m^{A_3} *are different:*
- $m^{A_1A_3\downarrow A_3}$ $(\{a_{31}\}) = 0.28$ *and* $m^{A_1A_3\downarrow A_3}$ $(\{a_{32}\}) = 0.72$ *if the path* A_1, A_3 *is followed.*
- $m^{A_2A_3\downarrow A_3}$ $(\{a_{31}\}) = 0.46$ *and* $m^{A_2A_3\downarrow A_3}$ $(\{a_{32}\}) = 0.54$ *if we choose to follow* A_2, A_3.

With belief function networks where conditionals are defined per single parent, it is possible to find several distributions for a given node according to the path followed in the network.

Thus, it does not guarantee the uniqueness of the solution.

Besides, since the fusion process is performed, we cannot reconstruct the local distribution from the global joint distribution. For these reasons, belief function network with individual parents and where focal elements are singletons cannot be considered as a generalization of Bayesian networks. However, note that when a Bayesian network admits a solution, it is the case also for *ENC*s and *DEVN*s.

A *DEVN* can be represented as a Bayesian network where conditional distributions specified per each edge are aggregated into a single conditional value. Nevertheless, this network remains nonequivalent to a Bayesian network.

3.4 A new belief network (BNC)

We have shown in the last section that the *DEVN* representation is a more efficient storage environment than the *EN*. Besides, it is more flexible than *ENC* since conditional distributions may be defined given all the parents without a need to have binary relations between nodes (i.e., merge parent nodes into one single parent).

In this section, we propose a new associational belief network. Close to *DEVN*, this network is called belief network with conditional beliefs denoted by *BNC*. It will serve as a basis for modeling causality. In this network beliefs are expressed in terms of basic assignments. The representation we propose is more flexible than *DEVN* since the conditional distributions can be defined in the context of one parent or for more than one node without necessarily have to define them for all the parents.
Remember that the computation of the joint distribution in the *DEVN* is based on the ballooning extension and the vacuous extension. We will show that these concepts unfairly share knowledge between subsets. Accordingly, our network will be based on new definitions (Boukhris, Benferhat, & Elouedi, 2011b) of these notions.

3.4.1 Definition of the belief network with conditional beliefs

A belief network with conditional beliefs (*BNC*) is a graphical model denoted $\mathcal{G}$. It is defined on two levels:

- Qualitative level: represented by a directed acyclic graph $\mathcal{G}=(V, E)$ where V is a set of nodes denoting the domain variables and E is the set of directed edges encoding the dependencies among variables. Each variable A_i is associated with a finite set namely its frame of discernment Θ_{A_i} representing all its possible instances, i.e., $\{a_{ij,j=1,\ldots,|\Theta_{A_i}|}\}$. A variable A_j is called a parent of a variable A_i if there is an edge pointing from A_j towards A_i. The set of all parents of A_i is denoted by $U(A_i)$. A subset from $U(A_i)$ is denoted by $PA(A_i)$ representing some parent nodes of A_i where a single parent is denoted by $PA_j(A_i)$. An instance from $U(A_i)$, $PA(A_i)$ or $PA_j(A_i)$ is denoted respectively by $u(A_i)$, $Pa(A_i)$ and $Pa_j(A_i)$.

Example 3.5. *Let us consider the network of Figure 3.6.*

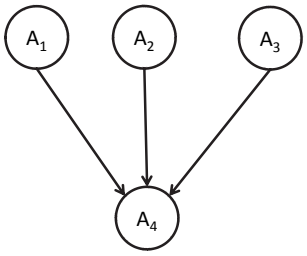


Figure 3.6: A belief network: BNC

- *The set of all parents of A_4 are denoted by $U(A_4) = \{A_1, A_2, A_3\}$.*
- *A single parent of A_4 is denoted by $PA_j(A_4)$ (i.e., $PA_1(A_4) = A_1$, $PA_2(A_4) = A_2$, $PA_3(A_4) = A_3$).*
- *A subset from $U(A_4)$ is denoted by $PA(A_4)$ (e.g., $\{A_1, A_2\}$, $\{A_3\}$, $\{A_1, A_2, A_3\}$) representing some or all parent nodes of A_4.*

- Quantitative level: represented by the set of *bba*s associated to each node in the graph. In this network, we work under the closed world assumption. Thus, the mass allocated to the empty set is equal to zero.

  - For each root node A_i (i.e., $PA(A_i) = \emptyset$) having a frame of discernment Θ_{A_i}, a normalized a priori m^{A_i} is defined on the powerset of $2^{\Theta_{A_i}}$, such that:

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}) = 1, \quad k = 2, \ldots, |2^{\Theta_{A_i}}|$$

    Besides, it is possible to model the total ignorance of the a priori by defining a vacuous *bba* on A_i (i.e., setting $\mathrm{m}(\Theta_{A_i}) = 1$).

  - For the rest of the nodes, conditional distributions can be defined for each subset of each variable A_i in the context of its parents (either one or more than one parent node): $m^{A_i}(.|Pa(A_i))$ such that:

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}|Pa(A_i)) = 1$$

3.4.2 Ballooning extension: from analysis to revision

Ballooning extension: why?

The term ballooning extension (Ristic & Smets, 2005; Ben Yaghlane & Mellouli, 2008) is used when for a given variable some new instances are added or discovered after that beliefs were already expressed by the experts. It is also used to get rid from conditioning as shown in Chapter 1. Therefore, the agent should redistribute his beliefs.

Ballooning extension: how?

Let us consider the case of beliefs defined on a frame of discernment Θ_X: m^X. From this *bba*, we would like to build a *bba* on a larger frame $\Theta_{X'} \supseteq \Theta_X$. This issue is of the domain of the ballooning extension: to compute $m^{X\uparrow X'}$. It amounts to transfer masses to the larger frame $\Theta_{X'}$. Such transfer can be done by three ways (Boukhris, Elouedi, & Benferhat, 2011a).

In the following, let us consider C a subset of Θ_X and A a subset of $\Theta_{X'}$. Let us denote by $\Theta_{\overline{X}}$: $\Theta_{X'}$-Θ_X.

- **Conservative transfer:** Intuitively one may consider that the *bbm* $m^{X\uparrow X'}(A)$ should be equal to the initial *bbm* defined on Θ_X. However, if this solution is retained the, masses defined on $2^{\Theta_{\overline{X}}}$ will be supposed to be always equal to zero since $\sum m^X(C)$ is already equal to one. Accordingly, this solution cannot be chosen. In fact, masses are not suitably transferred on the subsets. It assumed that the arrival of the new piece of information has no impact on the initial distribution.

$$m^{X\uparrow X'}(A) = \begin{cases} m^X(C) & \text{if } A = C \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

- **Least committed ballooning extension:** It is the solution proposed in (Smets, 1978). The new *bba* is built on the largest subset of $\Theta_{X'}$ that its intersection with Θ_X is Θ_X. Accordingly, original focal elements will no longer be considered as such. It is defined as:

$$m^{X\uparrow X'}(A) = \begin{cases} m^X(C) & \text{if } A = C \cup \Theta_{\overline{X}} \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

  Here also, masses are not suitably transferred on the subsets. In fact, the *bba* of the initial distribution is set to zero as we no longer believe it.

- **Uniform ballooning extension:** We propose a new definition denoted by $\Uparrow$. It allows the *uniform* transfer of masses. As for the least committed solution, the new *bba* is built on a subset of $\Theta_{X'}$ whose intersection with Θ_X is Θ_X. Since we are not only interested in the largest subset of $\Theta_{X'}$, several subsets satisfy this condition. Hence, the mass originally allocated to C will be reallocated many times to these different subsets. By this way, the sum of $m^{X\uparrow X'}(A)$ will also exceed 1. To tackle this problem, we propose to share uniformly the original *bbm* between all its supersets that contain the alternatives that were not taken into consideration. Formally:

$$m^{X\Uparrow X'}(A) = \frac{m^X(C)}{|2^{\Theta_{\overline{X}}}|}, C \subseteq A \tag{3.8}$$

The last solution that we propose is a good compromise between the two first alternatives. It allows to uniformly distribute initial beliefs. Accordingly, a non empty mass is attributed to the initial distribution and at the same time the new instances are taken into account.

Example 3.6. *A murder has been committed. There are two suspects* $\Theta_{person} = \{Bob\ (b), Sylvia\ (s)\}$. *Witnesses may express their beliefs about the murderer by declaring if they thought he is guilty or innocent (*$\Theta_{state} = \{Guilty\ (g), Innocent\ (i)\}$*).*

Their beliefs will be formalized on $\Theta_X = \Theta_{person} \times \Theta_{state}$.

Let us consider a subset C *defined on* Θ_X *such that* $C = \{(b,g), (s,g)\}$.
A witness defines his beliefs as follows:
$m^X(C)=0.7$, $m^X(\{(b,g)\}) = 0.1$, $m^X(\{(s,i)\}) = 0.2$.

If after that beliefs were defined on Θ_X, *we discover that there is another suspect John (j) (i.e.,* $\{(j,g), (j,i)\}$*). Beliefs defined on* Θ_X *should be reallocated to take into account this new information.*

Let $\Theta_{X'}$ *be a larger frame defined as* $\Theta_{X'} = \Theta_X \bigcup \{(j,g), (j,i)\}$.
We denote by $\Theta_{\overline{X}}$ *elements of* $\Theta_{X'}$-Θ_X *(i.e.* $\{(j,g), (j,i)\}$*).*
Let A *and* B *two subsets of* $\Theta_{X'}$ *composed of* C *and subsets of* $\Theta_{\overline{X}}$ *such that* $A = \{(b,g), (s,g), (j,i)\}$ *and* $B = \{(b,g), (s,g), (j,g)\}$*), then the possible reallocations are:*

- *conservative transfer:*
  $m^{X'}(A) = m^{X'}(B) = 0$;
  $m^{X'}(C)=0.7$; $m^{X'}(\{(b,g)\}) = 0.1$; $m^{X'}(\{(s,i)\}) = 0.2$.

- *least committed transfer:*
  $m^{X'}(A) = m^{X'}(B) = m^{X'}(C) = m^{X'}(\{(b,g)\}) = m^{X'}(\{(s,i)\}) = 0$;
  $m^{X'}(C \cup \Theta_{\overline{X}}) = 0.7$; $m^{X'}(\{(b,g)\} \cup \Theta_{\overline{X}}) = 0.1$; $m^{X'}(\{(s,i)\} \cup \Theta_{\overline{X}}) = 0.2$.

- *Uniform transfer: the bbm allocated to* C *will be transferred from* C *to* A *and also from* C *to* B. *Thus, if* $m^{X'}(B) = m^X(C)$ *and* $m^{X'}(A) = m^X(C)$ *then* $\sum_{A \subseteq \Theta_{X'}} m^{X'}(A) > 1$.
  Our proposed solution is to divide the bbm of C *equally between all subsets of* $\Theta_{\overline{X}}$. *Hence:*
  $m^{X'}(A) = m^{X'}(B) = m^{X'}(C) = m^{X'}(C \cup \Theta_{\overline{X}}) = 0.7/4 = 0.175$.

Uniform deconditioning on the product space

As an alternative to the conditional embedding (Dubois & Denœux, 2010; Smets, 1978), we propose a new definition for the deconditioning process that is based on a uniform transfer of basic assignments. Thus, not only least committed subsets will be considered as focal elements. This operation will be useful to transform a conditional *bba* in the *BNC* to a non-conditional one. The deconditioning operator will be denoted by $\Uparrow$.

Let $m^\Theta(. \mid \omega)$ be defined on Θ for $\omega \in \Omega$ and we want to get rid of conditioning. The idea is to divide the original mass defined on a given subset of Θ equally between all its supersets that contain the new discovered alternatives. Disposing of a conditioned *bba*, $m^\Theta(A \mid \omega)$, these alternatives are those in $\overline{\omega}$. Accordingly, the transfer of masses is computed as follows (Boukhris, Benferhat, & Elouedi, 2011b):

$$m_\omega^{\Theta \Uparrow \Theta\Omega}(C) = \begin{cases} \frac{m^\Theta(A|\omega)}{2^{|\Theta| \cdot |\overline{\omega}|}} & \text{if } C = (A \times \{\omega\} \cup B \times \overline{\omega}), B \subseteq \Theta \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

Example 3.7. *Let us take the same example presented in Chapter 1 for the definition of the state-of-the-art ballooning extension concept (Example 1.7). Initially, we have a conditional bba defined on* Θ=$\{Mary, John, Peter\}$ *in context of* $\Omega = \{Tall, Short\}$.

Let us assume that the witness has declared that the murderer is $Tall$. *The confidence that the murderer is Peter is defined with* $m^{\Theta}(Peter|Tall)$=0.7. *To get rid of conditioning and define this bbm on* $\Theta \times \Omega$, *we have to uniformly transfer the bbm allocated to* $\{(Peter, Tall)\}$ *to all subsets of* $\Theta \times \Omega$ *whose intersection with* $\{(Peter, Tall)\}$ *is not empty. Accordingly,*
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall)\})$ = 0.7/8 = 0.0875
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Mary,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Peter,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(John,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Mary,Short),(Peter,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Mary,Short),(John,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Peter,Short), (John,Short)\})$ = 0.0875.
$m^{\Theta \Uparrow \Theta\Omega}(\{(Peter,Tall),(Mary,Short),(Peter,Short), (John,Short)\})$ = 0.0875.

3.4.3 Vacuous extension: from analysis to revision

Vacuous extension: why?

The vacuous extension is used when the referential is changed, i.e., new variables are added. This notion is useful within belief function networks since it allows to define local distributions on the same frame of discernment in order to combine them afterwards. This operation will be useful to define a local *bba* in the *BNC* to a joint frame.

Vacuous extension: how?

In Chapter 1, we have seen that the mass allocated to subsets of the original variables will be transferred in a least committed way to all instances of the new variable. We propose a new definition where beliefs are uniformly shared between all subsets of the new variables. The operator will be denoted by $\Uparrow$. Note that this does not affect the definition of the cylindrical extension presented in Chapter 1 (see Equation 1.32).

Thus, a marginal mass function m^{Θ} defined on Θ will be normalized and expressed on the frame $\Theta \times \Omega$ as follows:

$$m^{\Theta \Uparrow \Theta\Omega}(C) = \begin{cases} \dfrac{m^{\Theta}(A)}{|2^{\Omega}| - 1} & \text{if } C^{\downarrow\Theta} = A \\ 0 & \text{otherwise} \end{cases} \quad , A \subseteq \Theta, C \subseteq \Theta \times \Omega. \tag{3.10}$$

Example 3.8. *Let us take the same example presented in Chapter 1 for the definition of the state-of-the-art vacuous extension concept (Example 1.5). Consider a bba defined on* $\Theta = \{Mary, John, Peter\}$ *as:* $m^{\Theta}(\{Mary\})$=0.5, $m^{\Theta}(\{John\})$=0.2, $m^{\Theta}(\{\Theta\})$=0.3.

The height of the murder is defined on $\Omega = \{Tall, Short\}$.

The bba defined on Θ *will be defined in a finer frame* $\Omega \times \Theta$ *using the uniform vacuous extension. For instance the bbm* $m^{\Theta}(\{Mary\})$=0.5 *will be extended as follows:*
$m^{\Theta \Uparrow \Theta\Omega}(\{(Mary, Tall)\})$= 0.5/3,

$m^{\Theta \nearrow \Theta\Omega}(\{(Mary, Short)\}) = 0.5/3$,
$m^{\Theta \nearrow \Theta\Omega}(\{(Mary, Tall), (Mary, Short)\}) = 0.5/3$.

3.4.4 Global joint distribution computation on *BNC*

The global joint distribution of the *BNC* is based on the uniform deconditioning and the uniform vacuous extension. Indeed, each local distribution should be first extended to a joint frame. Thus, each conditional distribution will be uniformly deconditionalized (denoted by $\Uparrow$) and non-conditionalized distribution will be uniformly vacuously extended to a joint frame (denoted by $\nearrow$). To compute the global joint distribution on the BNC, these three steps should be followed:

1. Getting rid from conditioning for each conditional variable A_i:

   1.1 For each instance of the parents of A_i denoted by $Pa(A_i)$ (defined for one or more than one parent node), compute the uniform deconditioning of $m^{A_i}(.|Pa(A_i))$:

   $$m^{A_i}(.|Pa(A_i))^{\Uparrow A_i PA(A_i)}$$

   1.2 Combine the deconditionalized *bba*s using Dempster's rule of combination.

   $$\bigoplus_{Pa(A_i)} m^{A_i}(.|Pa(A_i))^{\Uparrow A_i PA(A_i)}$$

2. Extend each node (root node and child node) to the universe of all variables in the network by applying the uniform vacuous extension.

   $$\left(\bigoplus_{Pa(A_i)} m^{A_i}(.|Pa(A_i))^{\Uparrow A_i \times PA(A_i)}\right)^{\nearrow A_{i=1,\ldots,n}}$$

3. Combine local joint distributions using Dempster's rule of combination.

   $$m^{A_1,\ldots,A_n} = \bigoplus_{i=1,\ldots,n} \left(\bigoplus_{Pa(A_i)} m^{A_i}(.|Pa(A_i))^{\Uparrow A_i PA(A_i)}\right)^{\nearrow A_{i=1,\ldots,n}} \quad (3.11)$$

Example 3.9. *Let us consider the belief network with conditional beliefs depicted in Figure 3.7 where A and C are binary variables such that* $\Theta_A = \{a_1, a_2\}$ *and* $\Theta_C = \{c_1, c_2\}$. *To compute the joint distribution, we propose to use the uniform deconditioning and vacuous extension. For that, we have to:*

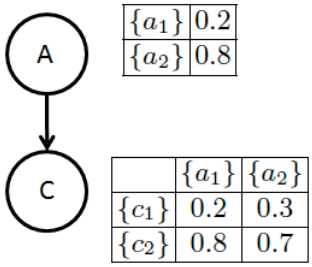


Figure 3.7: *BNC* with Bayesian *bba*

1. *Uniformly vacuously extend the bba m^A (see Table 3.11)*

Table 3.11: Uniform vacuous extension: $m^{A\nearrow AC}$

$m^{A\nearrow AC}(a_1)$	$\{(a_1, c_1)\}$=0.2/3
	$\{(a_1, c_2)\}$=0.2/3
	$\{a_1 \times \Theta_C\}$=0.2/3
$m^{A\nearrow AC}(a_2)$	$\{(a_2, c_1)\}$ =0.8/3
	$\{(a_2, c_2)\}$=0.2/3
	$\{a_2 \times \Theta_C\}$=0.2/3

2. *Get rid from conditioning using the uniform deconditioning: $m^C(.|a)^{\Uparrow AC}$ (see Table 3.12)*

Table 3.12: Uniform deconditioning: $m^C(.|a)^{\Uparrow AC}$

$m_{a_1}^{C\Uparrow AC}$		$m_{a_2}^{C\Uparrow AC}$	
$\{(a_1, c_1)\}$	0.05	$\{(a_2, c_1)\}$	0.075
$\{(a_1, c_1), (a_2, c_1)\}$	0.05	$\{(a_2, c_1), (a_1, c_1)\}$	0.075
$\{(a_1, c_1), (a_2, c_2)\}$	0.05	$\{(a_2, c_1), (a_1, c_2)\}$	0.075
$\{(a_1, c_1), (a_2, c_1), (a_2, c_2)\}$	0.05	$\{(a_2, c_1), (a_1, c_1), (a_1, c_2)\}$	0.075
$\{(a_1, c_2)\}$	0.2	$\{(a_2, c_2)\}$	0.175
$\{(a_1, c_2), (a_2, c_1)\}$	0.2	$\{(a_2, c_2), (a_1, c_1)\}$	0.175
$\{(a_1, c_2), (a_2, c_2)\}$	0.2	$\{(a_2, c_2), (a_1, c_2)\}$	0.175
$\{(a_1, c_2), (a_2, c_1), (c_1, a_2)\}$	0.2	$\{(a_2, c_2), (a_1, c_1), (a_1, c_2)\}$	0.175

3. *Combine local deconditionalized distribution of the variable C (see Table 3.13).*

Table 3.13: Local deconditionalized distribution

		0.075		0.075		0.075		0.075		0.175		0.175		0.175		0.175	
		{(a2,c1)}		{(a2,c1),(a1,c1)}		{(a2,c1),(a1,c2)}		{(a2,c1),(a1,c1),(a1,c2)}		{(a2,c2)}		{(a2,c2),(a1,c1)}		{(a2,c2),(a1,c2)}		{(a2,c2),(a1,c1),(a1,c2)}	
{(a1,c1)}	0.05	∅	0.00375	{(a1,c1)}	0.00375	∅	0.00375	{(a1,c1)}	0.00375	∅	0.0088	{(a1,c1)}	0.0088	∅	0.0088	{(a1,c1)}	0.0088
{(a1,c1),(a2,c1)}	0.05	{(a2,c1)}	0.00375	{(a2,c1),(a1,c1)}	0.00375	{(a2,c1)}	0.00375	{(a1,c1),(a2,c1)}	0.00375	∅	0.0088	{(a1,c1)}	0.0088	∅	0.0088	{(a1,c1)}	0.0088
{(a1,c1),(a2,c2)}	0.05	∅	0.00375	{(a1,c1)	0.00375	∅	0.00375	{(a1,c1)}	0.00375	{(a2,c2)}	0.0088	{(a1,c1),(a2,c2)}	0.0088	{(a2,c2)}	0.0088	{(a1,c1),(a2,c2)}	0.0088
{(a1,c1),(a2,c1),(a2,c2)}	0.05	{(a2,c1)}	0.00375	{(a2,c1),(a1,c1)}	0.00375	{(a2,c1)}	0.00375	{(a2,c1),(a1,c1)}	0.00375	{(a2,c2)}	0.0088	{(a1,c1),(a2,c2)}	0.0088	{(a2,c2)}	0.0088	{(a1,c1),(a2,c2)}	0.0088
{(a1,c2}	0.2	∅	0.015	∅	0.015	{(a1,c2)}	0.015	{(a1,c2)}	0.015	∅	0.035	∅	0.035	{(a1,c2)}	0.035	{(a1,c2)}	0.035
{(a1,c2),(a2,c1)}	0.2	{(a2,c1)}	0.015	{(a2,c1)}	0.015	{(a1,c2),(a2,c1)}	0.015	{(a1,c2),(a2,c1)}	0.015	∅	0.035	∅	0.035	{(a1,c2)}	0.035	{(a1,c2)}	0.035
{(a1,c2),(a2,c2)}	0.2	∅	0.015	∅	0.015	{(a1,c2)}	0.015	{(a1,c2)}	0.015	{(a2,c2)}	0.035	{(a2,c2)}	0.035	{(a1,c2),(a2,c2)}	0.035	{(a1,c2),(a2,c2)}	0.035
{(a1,c2),(a2,c1),(a2,c2)}	0.2	{(a2,c1)}	0.015	{(a2,c1)}	0.015	{(a1,c2),(a2,c1)}	0.015	{(a1,c2),(a2,c1)}	0.015	{(a2,c2)}	0.035	{(a2,c2)}	0.035	{(a1,c2),(a2,c2)}	0.035	{(a1,c2),(a2,c2)}	0.035

Table 3.14: Combination of $m^{A\curlywedge AC}$ with normalized $m^{C\Uparrow AC}$

mC(. | a) deconditionalized \ Vacuous extension of mA		0.06666667		0.06666667		0.06666667		0.26666667		0.26666667		0.26666667	
		{(a1,c1)}		{(a1,c2)}		{(a1,c1),(a1,c2)}		{(a2,c1)}		{(a2,c2)}		{(a2,c1),(a2,c2)}	
{(a1,c1)}	0.06666667	{(a1,c1)}	0.01	∅	0	{(a1,c1)}	0.01	∅	0	∅	0	∅	0
{(a2,t1)}	0.1	∅	0	∅	0	∅	0	{(a1,c2)}	0.06	∅	0	{(a2,c1)}	0.06
{(a1,c2)}	0.26666667	∅	0	{(a1,c2)}	0.04	{(a1,c2)}	0.04	∅	0	∅	0	∅	0
{(a2,c2)}	0.23333333	∅	0	∅	0	∅	0	∅	0	{(a2,c2)}	0.14	{(a2,c2)}	0.14
{(a1,c1),(a2,c1)}	0.02	{(a1,c1)}	0.003	∅	0	{(a1,c1)}	0.003	{(a1,c2)}	0.012	∅	0	{(a2,c1)}	0.012
{(a2,c1),(a1,c2)}	0.08	∅	0	{(a1,c2)}	0.012	{(a1,c2)}	0.012	{(a1,c2)}	0.048	∅	0	{(a2,c1)}	0.048
{(a1,c1),(a2,c2)}	0.04666667	{(a1,c1)}	0.007	∅	0	{(a1,c1)}	0.007	∅	0	{(a2,c2)}	0.028	{(a2,c2)}	0.028
{(a1,c2),(a2,c2)}	0.18666667	∅	0	{(a1,c2)}	0.028	{(a1,c2)}	0.028	∅	0	{(a2,c2)}	0.112	{(a2,c2)}	0.112

4. *Aggregate normalized* $m^{A\uparrow AC}$ *with normalized* $m^{C\Uparrow AC}$ *to get the joint mass distribution on* $\Theta_A \times \Theta_C$ *(see Table 3.14). Note that normalized results found on the* BNC *when bbas are Bayesian are the same as those found on Bayesian networks (see Table 3.15).*

Table 3.15: Normalized global joint distribution with uniform transfers

m^{AC}	
$\{(a_1, c_1)\}$	0.04
$\{(a_1, c_2)\}$	0.16
$\{(a_2, c_1)\}$	0.24
$\{(a_2, c_2)\}$	0.56

3.5 Conclusion

Belief networks are compact and flexible graphical representations. In this chapter, we have analyzed existing belief directed acyclic graphs. First, we have shown that the chain rule in Evidential networks *(EN)* is made by factorizing masses and have explained that mass functions are not not monotonous with respect to set inclusion. Then, we have shown that Evidential Networks with Conditional beliefs *(ENC)* and Directed EVidential Networks *(DEVN)* are not true extensions of Bayesian networks. Besides, they are based on the least committed vacuous and ballooning extensions.

We have proposed a new associational belief network *(BNC)* where conditional distributions are defined in terms of basic assignments defined for one or more than one parent node. Accordingly, it will have the advantage to simplify knowledge acquisition. It is based on fair transfers of masses namely the uniform vacuous extension and the uniform deconditioning. This network will serve as a base to the proposed causal belief network that will be introduced in the following chapter.

Chapter 4

Causal belief network with conditional beliefs

4.1 Introduction

Our proposed belief network called *BNC* presented in Chapter 3 allows the prediction of changes occurring on the system after some observations. Those observations happen by themselves without any manipulation on the system.

While an observation is a new information about the value of a variable in a static world, an intervention, (Halpern & Pearl, 2005; Pearl, 2000), is the effect of an external action that forces a variable to have a specific value in a dynamic world. This distinction is somewhat similar to the one between belief revision (Gänderfors, 1992) and update (Katsuno & Mendelzon, 1991) used for modeling belief change. However, in belief update there is no distinction between external and internal actions.

Since external actions affect the system differently, the reasoning process requires different modeling tools. It is important to note that despite their importance, no models handling interventions are provided under the belief function framework.

In this chapter, we present a causal model under the belief function framework namely causal belief network with conditional beliefs *(CBNC)*. It is an alternative to causal Bayesian network representation. The advantage of this network comparing to the Bayesian one, is that it allows the description of uncertain effects including situations of total ignorance after making an intervention.

Besides, we propose a counterpart of the "do" operator to the belief function framework. It will be useful to distinguish between observations and interventions. Graphically, external actions are handled through making changes on the structure of the graph. From reasoning point of view, after acting on a variable we assume that its initial causes are no more responsible of its state. Accordingly, arcs linking the variable of interest to its parents should be deleted. The resulting graph is a mutilated causal belief network. Another alternative is to add a new fictive variable DO as a parent node of the variable concerned by an intervention.

This added variable will totally control its status. The resulting graph is called an augmented causal belief network. Even if the computation of the belief joint distribution is different from the one used in probability theory, we show that the graph augmentation and graph mutilation are equivalent methods and lead to the same joint distribution.

However, in real world applications, assuming that an intervention is always perfect is not usually true. In fact, acting on a variable and setting it into one specific value is not always possible to achieve or ethic to do. Therefore, handling non-standard interventions is required. Despite its need in real world applications, only recent few works in the probabilistic setting addressed this issue (e.g., (Eberhardt & Scheines, 2007; Korb et al., 2004; Teng, 2012; Woodward, 2003)). In these works, interventions are considered as external actions represented with dummy variables that change the local probability distribution of the target variable. In our modeling, we consider the case of imperfect interventions that may have imperfect consequences.

The chapter is organized as follows: In Section 4.2, we propose a definition of causation and explain the difference between observations and interventions under a belief function framework. Then, in Section 4.3 we define causal belief networks that rely on belief function networks introduced in Chapter 3 and detail the changes on the graphs toward handling simple interventions. Section 4.4 deals with imperfect interventions while Section 4.5 presents the idea of interventions that may be imperfect and have imperfect consequences. Proofs of propositions provided in this chapter are presented in Appendix *A*.

4.2 Belief causation: observations vs interventions

Unlike deterministic approaches where causes are necessary to the occurrence of their effects, a belief causal link defines a higher belief of effects when a cause takes place and accordingly if a cause does not arise then the belief of the effect will decrease. External actions on the system disturb the relationships between variables and thus should have a different impact on the other events. The application of Dempster's rule of conditioning is appropriate when an event occurs spontaneously (observation) and will lead to erroneous results when something forces the event to happen (intervention).

In the following, we present the difference between standard interventions and observations under the belief function framework.

4.2.1 Observation

Observation is seeing. It can provide some information about the statistical relations amongst events. When we have passively observed an event, we can reason backwards diagnostically to infer the causes of this event, or we can reason forward and predict future effects.

Example 4.1. *Suppose that you are in West Palm Beach and have gone to a cafe and you have ordered a cup of coffee. Given your beliefs about how coffee is prepared in this area. Your initial beliefs are:*

$$m_1(sugar = \{no\}) > m_1(sugar = \{yes,\ no\}) > m_1(sugar = \{yes\}).$$

This reflects the fact that you are more confident that there is no sugar in coffee, and even if you have some doubts {(yes, no)}, you think that it is less likely that the coffee already contains sugar.

After you have tasted the coffee, you notice that it is sweet. Therefore, you need to revise your beliefs based on this observation. It may be:

$$0.9 = m_2(sugar = \{yes\}) > m_2(sugar = \{yes, no\}) > m_2(sugar = \{no\}) = 0.$$

In fact, even if you have noted that the coffee is sweet and you are pretty sure that they have changed their way to prepare the coffee by adding some sugar inside it, you still have some doubts. In fact, you think that it may be the effect of having ate ten minutes ago a pork sausage that was very salty, that let you found the coffee sweet.

4.2.2 Intervention

A standard intervention is an external action to the system that forces a target variable to have a specific value. It allows the identification of elements in a sequence of events that are related in a causal way.

Example 4.2. *Assume that before tasting the coffee, your friend has added all the quantity of the sugar in the container into your cup of coffee. It is obvious that this action has an impact on the sweetness of the coffee. It forces it to take the value "sweet". Note that this intervention does not affect your initial beliefs regarding the initial preparation of the coffee.*

$$\begin{aligned} &m'_1(sugar = \{no\}) = m_1(sugar = \{no\}) > \\ &m'_1(sugar = \{yes, no\}) = m_1(sugar = \{yes, no\}) > \\ &m'_1(sugar = \{yes\}) = m_1(sugar = \{yes\}) \end{aligned}$$

Like for probability and possibility theories, interventions should be distinguished from observations under the belief function framework. Observations concern static worlds while interventions concern dynamic worlds. The difference between observations and interventions is somewhat similar to the difference between belief revision (Gänderfors, 1992) and updating (Katsuno & Mendelzon, 1991). Indeed, belief revision is the process by which an agent changes his beliefs about a static world in the light of new information where belief update is the process by which an agent keeps his beliefs up to date with an evolving world. However, in updating there is no distinction between internal and external actions which is not the case for interventions. In fact, an intervention is by definition an external manipulation to the system.

A mentioned in Chapter 2, manipulations on the system disturb the relationships between variables, they should have a different impact on the other events than observations. In fact, it consists in ignoring the effects of all the causes related to the variable of interest. Therefore, while the effect of an observation is computed by the conditional belief $m(.|a_{ij})$, the effect of an external manipulation needs a different computational tool, namely, the "do" operator. This latter was originally introduced in (Goldszmidt & Pearl, 1992) for the ordinal conditional functions of Spohn (1988) and proposed after that in (Pearl, 2000) under a probabilistic framework. Thus, an intervention on a variable A_i that forces it to take the specific value a_{ij} is denoted by $do(a_{ij})$. In (Boukhris, Elouedi, & Benferhat, 2011b), we propose a counterpart of this operator since it is important to have a belief function framework allowing to model interventions and infer causal belief reasoning.

The following section defines causal belief networks and details how to deal with interventions. Hence, we explain how to compute the effect of these external actions, i.e., $m(.|do(a_{ij}))$.

4.3 Causal belief networks

A belief network represents an efficient way to model dependencies between variables and to predict the effect of observations on the joint distribution of the variables. Causal reasoning can be intuitively and formally described with graphs (Benferhat & Smaoui, 2007b; Pearl, 2000). On these networks, it is possible to predict the effects of both observations and external actions on the system. Proofs of propositions given in this chapter can be found in Appendix A.

4.3.1 Definition

A causal belief network with conditional beliefs *(CBNC)* (Boukhris, Elouedi, & Benferhat, 2011b) is based on the belief network *(BNC)* presented in Chapter 3.

Several *BNC*s are Markov equivalent in the sense that they model the same global joint distribution. At most, one of them follows the causal process which is the causal belief network. The latter consists of:

- A graphical component: It is a DAG, where nodes represent variables and arcs describe cause-effect relations. Directed edges encode *causal* relationships among variables. Parents of a given variable are seen as its immediate cause, and accordingly the equivalence hypothesis of belief function networks is not valid any more. The set of parents of A_i is denoted by $PA(A_i)$.
- A numerical component: It is the set of normalized *bba*s associated to each node in the graph.
  - For each root node A_i (i.e., $PA(A_i) = \emptyset$) having a frame of discernment Θ_{A_i}, a normalized a priori m^{A_i} is defined on the powerset of $2^{\Theta_{A_i}}$, such that:

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}) = 1, k = 2, \ldots, |2^{\Theta_{A_i}}|$$

  - For the rest of the nodes, conditional distributions can be defined for each subset of each variable A_i in the context of its parents (either one or more than one parent node): $m^{A_i}(.|Pa(A_i))$ such that:

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}|Pa(A_i)) = 1$$

An external action (interventional data) will affect the system differently and consequently will lead to different results than those found with observational data. These effects should be adequately predicted.

As seen in Chapter 2 that under a probabilistic or possibilistic framework, handling interventions and computing their effects on the system can be done by making two different but

equivalent changes on the structure of the causal network: the so-called graph augmentation and graph mutilation methods.

In the following subsections, we will investigate these changes on the causal belief graph and will show that even if the belief joint distribution is not defined as for probability distribution, these methods are equivalent and lead to the same joint distribution under the belief function framework.

4.3.2 Intervention by graph mutilation

An intervention puts its target variable into exactly one specific value. Therefore, this action makes the direct causes (parents) of the variable concerned by the intervention not more responsible of its state. However, beliefs on these direct causes should not be modified.

Thus, all the edges directed to the target node will be deleted. No changes affect other nodes. Pearl (2000) considers it as a surgery (a mutilation) by which all the other causes than the one of the intervention will be excluded.

Definition 4.1. *Let $\mathcal{G}$ be a causal belief network on which we make an intervention on a variable $A_i \in V$, where V is the set of all the nodes and v is a subset from the cartesian product of variables in V. Intervening on A_i means that we force it to take the specific value a_{ij} ($do(a_{ij})$). We define mutilation on two steps:*

1. *Arcs pointing to A_i in $\mathcal{G}$ will be deleted. The obtained mutilated graph is denoted $\mathcal{G}_{mut}$. Its associated belief distribution is denoted $m_{\mathcal{G}_{mut}}$. The intervention $do(a_{ij})$, affects the computation of the joint distribution $m_{\mathcal{G}}$ by transforming it into $m_{\mathcal{G}}(.|do(a_{ij}))$. On the mutilated graph, it corresponds to observing a_{ij}. Thus, it simply consists of conditioning the mutilated graph by the value a_{ij}*

$$m^V_{\mathcal{G}}(.|do(a_{ij})) = m^V_{\mathcal{G}_{mut}}(.|a_{ij}) \tag{4.1}$$

2. *An action $do(a_{ij})$ imposes the value a_{ij} on the variable A_i. Accordingly, the corresponding bba of A_i becomes certain as follows:*

$$m^{A_i}(sub_{ik}) = \begin{cases} 1 & \text{if } sub_{ik} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

As stated in the following proposition, to perform a causal inference we could simply make a conditioning on the mutilated graph by the target value of the variable concerned by the intervention, namely a_{ij}.

Proposition 4.1. *Let $\mathcal{G}$ be a causal belief network and $m^V_{\mathcal{G}}$ be the joint mass distribution related to $\mathcal{G}$. The effect of an intervention $do(a_{ij})$ on the mass distribution is given by:*

$$m^V_{\mathcal{G}}(v|do(a_{ij})) = \begin{cases} \sum_{v' \cap a_{ij}^{\uparrow V} = v} m_{\mathcal{G}}^{V \setminus A_i}(v')^{\uparrow V} & \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where v is a subset from the cartesian product of the variables in V.

Example 4.3. *(continued)*

If your friend adds sugar into your cup of coffee then it will affect its sweetness. It is considered as an external intervention on the variable C that forces it to take the value "sweet" (i.e., $c = \{c_1\}$). Graphically, this action will lead to the disconnection of C from its original cause, here the initial use of sugar S. Initial beliefs regarding the initial use of sugar remain unchanged, i.e., m^S.

Figure 4.1 illustrates the new graph, the mutilated one where the parents of the manipulated variable become independent from it.

Note that the effects of observations on the graph of Figure 3.1 or on its non-causal Markov equivalent networks are the same. However, results will be different after intervening on some variables. Indeed, observing the value c_1 on the Markov equivalent networks will lead to the same results whereas manipulating the variable C will give different mutilated graphs and consequently different results.

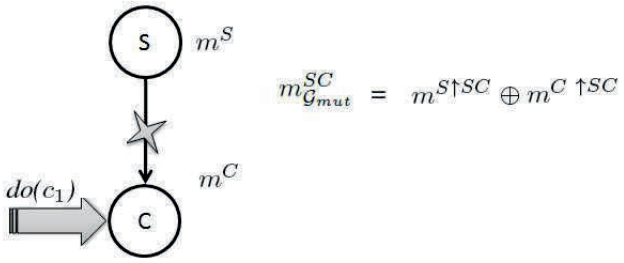


Figure 4.1: Mutilated causal belief graph

Observing the value c_1 on the mutilated graph $\mathcal{G}_{mut}$, represents the effect of the intervention $do(c_1)$ on the initial graph $\mathcal{G}$. It is computed by conditioning the joint mass distribution $m^{SC}_{\mathcal{G}_{mut}}$ by c_1. In fact, upon intervening on C, the event c_1 becomes certain. $m^C(c)$, for each $c \subseteq \Theta_C$, is computed by the application of Equation 4.2.

$$m^C_{\mathcal{G}_{mut}}(c) = \begin{cases} 1 & \text{if } c = \{c_1\} \\ 0 & \text{otherwise} \end{cases}$$

As S and C become independent after the intervention, their combination can be computed by making the pointwise product of their masses. Let us denote by s any subset of Θ_S and by c any subset of Θ_C. Then, $m^{SC}(s \times c)$ is computed as:
- $m^S(s) \cdot m^C(c) = m^S(s)$, if $(s \times c)^{\downarrow C} = \{c_1\}$.
- 0, otherwise.

Focal elements of m^{SC} are presented in Table 4.1.

Table 4.1: Focal elements of $m^{SC}_{\mathcal{G}_{mut}}$

$\{(s_1, c_1)\}$	0.1
$\{(s_2, c_1)\}$	0.7
$\{(s_1, c_1), (s_2, c_1)\}$	0.2

In the case where the initial *bba* of the target variable is non-dogmatic, the effect of the intervention can be also computed using the initial distribution. This is done by the application of the decombination operator to remove the conditional mass allocated to the

target variable.

Proposition 4.2. *Let $\mathcal{G}$ be a belief causal network in which a variable A_i is forced it to take the value a_{ij}, $do(a_{ij})$. In the case where the initial bba of the target variable is non-dogmatic, the effect of this intervention on the joint mass distribution $m^V_{\mathcal{G}}$ is given as follows:*

$$m^V_{\mathcal{G}}(v|do(a_{ij})) = \begin{cases} \sum\limits_{\mathcal{F}\cap a_{ij}^{\uparrow V}=v} m^V_{\mathcal{G}} \ominus (\bigoplus\limits_{Pa(A_j)} m^{A_i}_{\mathcal{G}}(.|Pa(A_i))^{\uparrow A_j \times PA(A_j)})^{\uparrow V} & \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

where $\mathcal{F}$ represent focal elements of $(m^V_{\mathcal{G}} \ominus (\bigoplus\limits_{Pa(A_j)} m^{A_i}_{\mathcal{G}}(.|Pa(A_i))^{\uparrow A_j \times PA(A_j)})^{\uparrow V})$ and v any subset of the cartesian product of variables in V.

Corollary 4.1. *Let $\mathcal{G}$ be a belief causal network whose joint mass distribution is $m^V_{\mathcal{G}}$. In the case where the initial bba of the target variable is non-dogmatic, the effect of an intervention $do(a_{ij})$ on a variable A_i of this graph forcing it to take the value a_{ij} can be also computed as follows:*

$$m^V_{\mathcal{G}}(v|do(a_{ij})) = \begin{cases} \sum\limits_{\mathcal{F}\cap\{(a_{ij},Pa(A_i))\}^{\uparrow V}=v} m^V_{\mathcal{G}}(.|a_{ij}, Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V} & \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

where $\mathcal{F}$ is a focal element of $m^V_{\mathcal{G}}(.|a_{ij}, Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V}$ and v is a subset from the cartesian product of variables in V.

4.3.3 Intervention by graph augmentation

Another possible interpretation of interventions on a causal belief network is to add a fictive node called "DO" as a parent of the target variable A_i (Pearl, 2000). Accordingly, the set of the parents of A_i is transformed from $U(A_i)$ to $U(A_i) \cup DO$. The augmented graph with the fictive nodes is denoted $\mathcal{G}_{aug}$. In this section, we will show that this representation is equivalent to the graph mutilation method to compute the effect of interventions.

The DO node is considered as an extra node in the system. It is taking value in $do(x)$, $x \in \{\Theta_{A_i} \cup \{nothing\}\}$.

$do(nothing)$ means that there are no actions on the variable A_i, it represents the state of the system when no interventions are made. $do(a_{ij})$ means that the variable A_i is forced to take the specific value a_{ij}.

The main advantage of handling intervention by graph augmentation, is that it allows to represent the effect of interventions and also observations which is not the case with graph mutilation method. In fact, by mutilating the graph it is not more possible to compute the effect of observations.

Let us first consider the changes on the *bba* of the target variable A_i. For each $sub_{ik} \subseteq \Theta_{A_i}$, the part of belief $m^{A_i}_{\mathcal{G}_{aug}}(sub_{ik}|Pa(A_i), do(x))$ is computed as follows:

$$m^{A_i}_{\mathcal{G}_{aug}}(sub_{ik}|Pa(A_i), do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik} = \{a_{ij}\} \\ 0 & \text{if } x \neq \{a_{ij}\} \\ m^{A_i}_{\mathcal{G}}(a_{ij}|Pa_j(A_i)) & \text{if } x = \{nothing\} \end{cases} \tag{4.6}$$

Regarding the *bba* assigned to the added fictive node (i.e., DO), two cases are considered:

- No interventions are performed. It allows to model the effect of observations as on the initial causal graph. $m^{DO}_{\mathcal{G}_{aug}}(do(x))$ is defined by:

$$m^{DO}_{\mathcal{G}_{aug}}(do(x)) = \begin{cases} 1 & \text{if } x = \{nothing\} \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

  The graph augmentation method allows to compute the effect of observations. As stated in the following proposition, it consists of conditioning the augmented graph by the value $do(nothing)$.

  Proposition 4.3. *Let $\mathcal{G}_{aug}$ be an augmented causal belief graph where the DO node is set to the value nothing.*
  Its corresponding bba $m^{V}_{\mathcal{G}_{aug}}(.|do(nothing))$ encodes the same joint distribution as the original causal belief graph.

$$m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^{V}_{\mathcal{G}} \tag{4.8}$$

  where $V' = V \cup DO$.

- If there is an intervention forcing the variable A_i to take the value a_{ij}, then the $m^{DO}_{\mathcal{G}_{aug}}(do(x))$ is defined by:

$$m^{DO}_{\mathcal{G}_{aug}}(do(x)) = \begin{cases} 1 & \text{if } x = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

  The graph augmentation method allows also to compute the effect of interventions by conditioning the augmented graph by the value $do(a_{ij})$. In this situation, belief graph mutilation and belief graph augmentation are equivalent methods for handling interventions.

  Proposition 4.4. *Let $\mathcal{G}$ be a causal belief network and $\mathcal{G}_{mut}$ and $\mathcal{G}_{aug}$ its corresponding mutilated and augmented graphs after acting on the variable A_i by forcing it to take the value a_{ij}. Then as for probability and possibility theories, computing the effects of interventions using the mutilation of the graph or its augmentation gives the same results.*

$$\begin{array}{ll} & m^{V}_{\mathcal{G}}(.|do(a_{ij})) \\ = & m^{V}_{\mathcal{G}_{mut}}(.|a_{ij}) \\ = & m^{V}_{\mathcal{G}_{aug}}(.|do(a_{ij})) \end{array} \tag{4.10}$$

We note that even though there is a difference in the construction of the global joint distribution between the initial, the mutilated and the augmented graph, the result remains the same. Given an initial causal belief network, acting on a given variable A_i by forcing it to take a specific value a_{ij} amounts to observing the value a_{ij} in its mutilated graph or observing the value $do(a_{ij})$ on the fictive node DO in its associated augmented network.

Example 4.4. *(continued)*

Acting on the variable C forcing it to take the value 'sweet' i.e., $do(c_1)$ is graphically represented by the augmented causal belief network shown in Figure 4.2. Aside from the direct cause S of the manipulated variable C, a new fictive node DO is added to the set of the parents of C where its frame of discernment is denoted by Θ_{DO}. A new distribution $m_{\mathcal{G}_{aug}}(s \times c \times \{do(x)\})$ has to be taken into consideration. It is defined on $\Theta_S \times \Theta_C \times \Theta_{DO}$. The new conditional bba on C given its new set of parents is shown in Table 4.2.

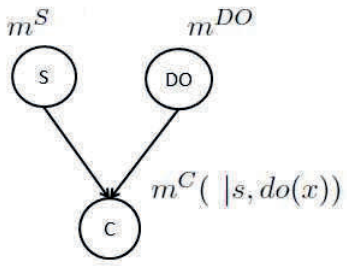


Figure 4.2: Augmented causal belief graph

Table 4.2: Conditional distribution $m^C(.|s, do(x))$

	$\{(s_1, do(nothing))\}$	$\{(s_1, do(c_1))\}$	$\{(s_1, do(c_2))\}$	$\{(s_2, do(nothing))\}$	$\{(s_2, do(c_1))\}$	$\{(s_2, do(c_2))\}$
$\{c_1\}$	0.8	1	0	0.1	1	0
$\{c_2\}$	0.1	0	1	0.7	0	1
Θ_C	0.1	0	0	0.2	0	0

In Table 4.3, we show the effect of an intervention $do(x) = do(c_1)$ on the distribution of the node DO (using Equation 4.9). The global joint distribution presented in Table 4.4 is given by the combination of the vacuous extension of m^S and m^{DO} with the ballooning extension of each $m^C(.|s, do(x))$.

Table 4.3: Local distribution m^{DO}

$do(nothing)$	0
$do(c_1)$	1
$do(c_2)$	0

Table 4.4: Focal elements of $m^{C \times S \times DO}$

$\{(c_1, s_1, \text{do}(c_1))\}$	0.1
$\{(c_1, s_2, \text{do}(c_1))\}$	0.7
$\{(c_1, s_1, \text{do}(c_1)),(c_1, s_2, \text{do}(c_1))\}$	0.2

Note that the conditioning of the distribution presented in Table 4.4 by $do(c_1)$ leads to the same results as those found in Table 4.1 resulting from a conditioning on the mutilated graph $m_{\mathcal{G}_{mut}}(.|c_1)$.

4.4 Dealing with imperfect interventions

Motivated by the fact that in real world applications forcing a variable to take one certain specific value is not generally feasible. We propose to handle non-standard interventions in our modeling.

4.4.1 Related works under a probabilistic framework

Despite their need in real world applications, only few works in the probabilistic setting addressed the issue of non-standard interventions (e.g., (Eberhardt & Scheines, 2007; Korb et al., 2004; Teng, 2012; Woodward, 2003)). They are considered as external actions that change the local probability distribution of the target variable. They are defined as:

- exogenous variables i.e., uncaused, with states reflecting whether the intervention is active or not.
- direct causes of target variables.

Let us consider that A_i is the manipulated variable. When the intervention is active ($i.e., on$), it is categorized as:

- independent (hard, structural, surgical): leads to an achieved distribution which is a function only for the new distribution aimed for the intervention. i.e., a new distribution P^* is defined where $P(A_i|Pa(A_i), on)$ is changed to $P^*(A_i)$.
- dependent (soft, parametric, conditional): leads to an achieved distribution which is a function from both the target distribution and the variable's causes (parents), i.e., a new distribution P^* is defined where $P(A_i|Pa(A_i), on)$ is transformed to $P^*(A_i|Pa(A_i))$.

It also varies by being either:

- deterministic: leaves the target variable into one state.
- stochastic: leaves the target variable with a new distribution with positive probability over two or more states.

Example 4.5. *(Independent stochastic intervention)*

Let us continue with the network of Figure 3.2. Suppose that your friend puts something into the cup of coffee. It is considered an external intervention on the variable C. Accordingly, the DO node in this case in set to the value "on". Graphically, this action will lead to the disconnection of C from its original cause. However, a stochastic distribution will be assigned to the node C. Indeed, the two states of the variable C (i.e., c_1 and c_2) will have positive probabilities. Table 4.5 represents the new probability distribution P^* of the target variable.

Table 4.5: Probability distribution of node C after an independent stochastic intervention

c_i	$P^*(c_i)$
c_1	0.7
c_2	0.3

Standard or ideal interventions are deterministic independent external actions. In the following subsections, we propose methods to handle non-standard interventions under the belief function framework. We will use these following qualifiers:

- *standard* interventions: external actions that succeed to force a variable to take a specific fixed and known value. A certain *bba* is specified for the target variable.
- *imperfect* interventions: experiments whose occurrence is imperfect (either *imprecise or uncertain*). In other words, imperfection concerns the DO node. These kinds of interventions have never been addressed before. In fact, in the probabilistic works the exogenous variable took a certain value.
- interventions with *imperfect consequences*: manipulations that may fail to put their targets into exactly one state. Therefore, like for probabilistic works (stochastic interventions) a new *bba* expressing the uncertainty is specified for the target variable. This *bba* can be defined from the interaction between the effect of the intervention and the initial causes of the target variable.

Accordingly, we will go further in the analysis of non-standard interventions. In fact, will explain how to define the new distribution of the DO node (i.e., the exogenous variable) as well as the one of the target variable. Note that imperfect interventions may or not have imperfect consequences.

4.4.2 Imprecise interventions

With standard interventions, the experimenter is assumed to completely control his manipulation. In fact, he succeeds to put the target variable into exactly one specific known value. Here, we consider another type of non-ideal interventions, namely those that act on one target variable and totally control its state as it is the case for standard interventions but the experimenter is not certain regarding the target values of his action. Note that this type of interventions cannot be handled under the probabilistic framework.

We choose to take advantage of the belief function theory to model such manipulations having not only certain but also imprecise target values. Motivational factors in this choice are:

- this task is not trivial within the probabilistic framework and this problem has never been addressed before.
- we only have a partial specification. In fact, we do not know precisely the effects of the imprecise intervention.
- the belief function framework provides a natural framework to represent such cases of ignorance. Indeed, it allows to express beliefs in terms of subsets instead of singletons. Imprecision is naturally represented within the belief function formalism with a *categorical bba* for the variable concerned by the intervention focused on a subset of its frame of discernment.

Remember that the "do" operator is used to distinguish interventions from observations. Upon an intervention, the original causes of the manipulated variable are no more responsible

of its state and all the other causes than the one of the intervention will be excluded. The experimenter puts the target variable A_i into exactly one specific state $a_{ij} \in \Theta_{A_i}$. Accordingly, the *bba* of the target variable becomes a certain *bba*.

In the following, we will expose how to deal with interventions on belief causal networks when the target value is imprecise, i.e., $do(sub_{ik}), sub_{ik} \subseteq \Theta_{A_i}$ and detail the changes according to standard interventions. We will show that graph mutilation and graph augmentation are adequate methods to model imprecise interventions and they lead to the same results.

Graph mutilation for imprecise interventions

The manipulated variable becomes uniquely determined by the external action, the intervention removes the influences of any other causes. Hence, the *bba* of the node A_i concerned by an intervention is no more as a conditional *bba* given its initial causes.

Imprecisely intervening means that the experimenter have some doubts concerning the target values of his manipulation. To model this kind of situations, a *bbm* of one is allocated to the subset composed by possible values of the target variable.

Let us denote by sub_{ik} the subset containing possible values that may take the variable A_i concerned by the intervention (i.e. $sub_{ik} \subseteq \Theta_{A_i}$). One alternative to compute the effect of this intervention is to mutilate the graph. It corresponds to alter the structure of the initial network as follows.

1. cut-off the arcs pointing to the target variable.
2. changes of the initial conditional *bba* to a categorical *bba* focused on sub_{ik} where:

$$m^{A_i}(sub_{ij}) = \begin{cases} 1 & \text{if } sub_{ij} = sub_{ik} \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

Proposition 4.5. *The global joint distribution obtained after intervening on the initial graph $\mathcal{G}$ by forcing a variable A_i to take the imprecise value sub_{ik} leads to the same results obtained after observing sub_{ik} on the mutilated graph $\mathcal{G}_{mut}$:*

$$m^V_{\mathcal{G}_{mut}}(.|sub_{ik}) = m^V_{\mathcal{G}}(.|do(sub_{ik})) \tag{4.12}$$

Graph augmentation for imprecise interventions

For graph augmentation method, a new node *DO* is added as a new parent of the node A_i on which an external action has been made. This method allows to compute the effect of not only interventions but also observations. In fact, the *DO* node is taking value in $do(x)$, $x \subseteq \{\Theta_{A_i} \cup \{\text{nothing}\}\}$. $do(nothing)$ means that there are no actions on the variable A_i, it represents the state of the system when no interventions are made. $do(sub_{ik}))$ means that the variable A_i is forced to take the imprecise value sub_{ik}.

When no interventions are made, the *DO* node is set to the value *nothing* and the *bba*s of the other variables remain unchanged. Experimentally, the agent will be passive. Indeed, he

observes the natural behavior of the system. Formally, the *bba* of the *DO* node is defined by:

$$m^{DO}_{\mathcal{G}_{aug}}(do(x)) = \begin{cases} 1 & \text{if } x = \{nothing\} \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

When an intervention is performed, the *bba* of the node *DO* is defined as a categorical *bba* where:

$$m^{DO}_{\mathcal{G}_{aug}}(do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik}, sub_{ik} \subseteq \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

For the *bba* of the target node A_i, namely $m^{A_i}_{\mathcal{G}_{aug}}(.|Pa(A_i), do(x))$, it becomes either a categorical *bba* when an intervention is performed or remains the same as on the initial graph in the case of no manipulations. It is defined as follows:

$$m^{A_i}_{\mathcal{G}_{aug}}(sub_{ik}|Pa(A_i), do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik} = sub_{ij} \\ 0 & \text{if } x \neq sub_{ik} \\ m^{A_i}_{\mathcal{G}}(sub_{ij}|Pa(A_i)) & \text{if } x = \{nothing\} \end{cases} \tag{4.15}$$

Proposition 4.6. *Let $\mathcal{G}$ a belief causal network and let $\mathcal{G}_{mut}$ and $\mathcal{G}_{aug}$ its corresponding mutilated and augmented graphs. Dealing with imprecise interventions using the mutilation of the graph or its augmentation gives the same results.*

$$\begin{array}{ll} & m^V_{\mathcal{G}}(.|do(sub_{ik})) \\ = & m^V_{\mathcal{G}_{mut}}(.|sub_{ik}) \\ = & m^V_{\mathcal{G}_{aug}}(.|do(sub_{ik})) \end{array} \tag{4.16}$$

Proposition 4.7. *Standard interventions are a particular case of imprecise interventions when the the subset representing the possible target values is composed of one element, i.e. $sub_{ij} = \{a_{ij}\}$.*

$$m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik} = \{a_{ij}\}, sub_{ik} \subseteq \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

Proposition 4.8. *Interventions with ignored target values are a particular case of imprecise interventions: the subset representing the possible target values corresponds to the frame of discernment, i.e. $sub_{ij} = \Theta_{A_i}$.*

$$m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik}, sub_{ik} = \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases} \tag{4.18}$$

Example 4.6. *(continued)*

Your friend sees on the table a container with some white powder. Without tasting it, he adds a large amount of this powder into your cup because he knows that you like sweet coffee. Later, he realizes it might be saccharin. In fact, many cafes offer saccharin instead of sugar because of the increasing number of diabetics and persons wishing to lose weight. However, even if saccharin has a sweetening power, it has a bitter unpleasant aftertaste at high concentrations. As he adds a large amount of this white powder, the coffee may become

bitter if it is saccharin.

Consequently, interventions are clearly external actions (of the system) that force variables to take some specific values. Here, it forces the coffee to take the value {sweet ,bitter} (i.e., $\{c_1, c_2\}$) reflecting the ignorance of the friend about the target values of his manipulation.

However, this intervention does not affect the initial beliefs regarding the initial use of sugar during the preparation of the coffee.

$$m_1''(sugar = \{no\}) = m_1(sugar = \{no\}) >$$
$$m_1''(sugar = \{yes, no\}) = m_1(sugar = \{yes, no\}) >$$
$$m_1''(sugar = \{yes\}) = m_1(sugar = \{yes\})$$

The external action of the friend can be described graphically in belief causal networks in two equivalent ways. The first way consists in the deletion of links from S (representing the initial use of sugar) pointing into C (representing the taste of the coffee). Therefore, the target variable becomes independent from its initial causes. The obtained network is called a mutilated belief network for imprecise interventions.

The intervention imposes on the variable concerned by the intervention, C, the value $\{c_1, c_2\}$. It reflects the ignorance about the taste of the coffee after acting on it (i.e., it may be sweet or bitter). Consequently, the conditional bba on C defined for $c \subseteq \Theta_C$ becomes a vacuous bba*:*

$$m^C(c) = \begin{cases} 1 & \text{if } c = \Theta_C \\ 0 & \text{otherwise} \end{cases}$$

The second equivalent way to graphically representing interventions consists in adding a new variable denoted DO as a new parent of the target variable C. The obtained network is called an augmented belief network for imprecise interventions. Unlike mutilated graphs, it is possible to compute on this network the effect of interventions and also observations. In fact, when the DO node is instantiated with the value nothing it allows to compute the effect of observational data.

The DO node will have more possible states compared to standard interventions (i.e., those reflecting imprecision about the target values, namely, subsets of Θ_C instead of only those belonging to the domain of C).

In the coffee example, the action of the friend is denoted by $do(\{c_1, c_2\})$ (or $do(\Theta_C)$). It has the following impact on the local distribution of the DO node.

$$\forall x \subseteq \Theta_C,\ m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = \Theta_C \\ 0 & \text{otherwise} \end{cases}$$

The bba of the target variable is defined as follows:

$$m^C(c|Pa(C), do(x)) = \begin{cases} 1 & \text{if } x = c = \Theta_C \\ 0 & \text{if } x \neq \Theta_C \\ m^C(c|Pa(C)) & \text{if } x = \{nothing\} \end{cases}$$

Consequently after the action of the friend, the conditional distribution of the target variable becomes a vacuous bba, $m^C(\Theta_C|Pa(C), do(\Theta_C)) = 1$.

Note that acting on the variable C by forcing it to take the value Θ_C in an initial belief causal network is equivalent to observing the value Θ_C in its associated mutilated graph or observing the value $do(\Theta_C)$ on the node DO in its corresponding augmented belief network.

4.4.3 Uncertain interventions

An intervention having the variable A_i as target may be uncertain. It means that it may uncertainly occur by forcing A_i to take *a specific unknown value* $a_{ij}(a_{ij} \in \Theta_{A_i})$ or *does not take place*. To represent such cases, we will add to the belief network the DO node. On the augmented graph, in a "natural" way a conditional *bba* in the context of the fictive node DO will be specified.

Interventions with an unknown specific value

In the following, we focus on the case when the interventions uncertainly happen. To compute the distribution of the target variable, we will detail four steps that should be considered.

1- Deciding about the nature of the external action

Unlike standard interventions where we are sure about the nature of the intervention, a normalized *bba* m^I expressing the beliefs about the genuine nature of the external action is defined on a frame of discernment $\Theta_I = \{\theta_1, \ldots, \theta_n\}$. Since an intervention is an intended action targeting a specific value of the variable concerned by the manipulation, deciding about the actual nature of the intervention will allow us to know which states will be affected by a change. The decision operation is made using the pignistic transformation. Hence, each induced $\text{BetP}(\theta_i)$ takes into account all focal elements of m^I intersecting with θ_i.

Example 4.7. *(continued)*
*Suppose that the beliefs about the nature of the substance in the container are flexibly expressed within the belief function formalism. They are defined on Θ_I=\{sugar, salt, heroine\} such that m^I(\{sugar\})=0.2, m^I(\{salt\})=0.7, m^I(\{heroine\})=0.01 and m^I(\{sugar,salt\})= 0.09. The corresponding probabilistic knowledge of this bba is computed with the pignistic probability measure as follows: BetP(\{sugar\})= 0.2+0.09*0.5=0.245, BetP(\{heroine\}) = 0.01, BetP(\{salt\})=0.7+0.09*0.5=0.745.*

2- Defining the possible states of the intervention

The frame Θ_I is different from the frame of the target variable Θ_{A_i}. However, instances of Θ_I may affect the state of the target variable A_i by forcing it to take the value a_{ij}. Thus in the case of uncertain interventions, a matching between each θ_i and a state from Θ_{A_i} is defined as $match(\theta_i) = a_{ij}$. If θ_i has no impact on A_i, then we will say that $match(\theta_i) = nothing$. Note that more than one element of Θ_I may affect the same state a_{ij}.

Example 4.8. *(continued)*
Let us continue with the last example where the target variable C has a frame of discernment Θ_C=\{c_1=sweet, c_2=bitter\} and the intervention has a Θ_I=\{sugar, salt, heroine\}. Table 4.6 presents the results of the matching between elements θ_i with instances of C.

Table 4.6: Matching function: $match(\theta_i)$

θ_i	$match(\theta_i)$
sugar	c_1
salt	nothing
heroine	c_2

Recall that the *DO* node represents the intervention. It has the same instances than its target to which the value *nothing* is added. $do(a_{ij})$ means that the intervention attempts to set the target variable A_i into the state a_{ij}. This is achieved by performing the action θ_i. Therefore, executing θ_i amounts to $do(a_{ij})$.

Accordingly, beliefs about the state of the variable *DO* reflecting the occurrence of the intervention will be defined from the knowledge about the decided nature of the intervention computed in the last step through BetPs. Since this latter reflects a probabilistic knowledge (i.e. computed for singletons), the *bba* of the *DO* node will be Bayesian and defined as:

$$m^{DO}(do(x)) = \begin{cases} \sum\limits_{\theta_i, match(\theta_i)=a_{ij}} BetP(\theta_i) & \text{if } x = \{a_{ij}\} \\ \sum\limits_{\theta_i, match(\theta_i)=nothing} BetP(\theta_i) & \text{if } x = \{nothing\} \end{cases} \tag{4.19}$$

Example 4.9. *(continued)*

According to the added substance, the coffee will be either sweet, bitter or remain as it was prepared. Therefore, forcing it to be at a specific state is not given for sure by adding the white powder. Hence, beliefs expressed about the actual occurrence of the intervention are computed using the BetP of each ingredient. In fact, the BetP takes into account all the focal elements that intersect with the substance of interest. The bba of the node DO is defined as: $m(\{do(c_1)\})$ *= BetP(sugar) = 0.245,* $m(\{do(c_2)\})$ *= BetP(heroine) = 0.01 and* $m(\{do(nothing)\})$ *= BetP(salt) = 0.745.*

2- Defining conditionals given the *DO* node

When occurring, an intervention $do(a_{ij})$ succeeds to force the variable A_i to take a certain value a_{ij}. Therefore, a conditional *bba* given an intervention is a certain *bba* focused on a_{ij} defined as:

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} 1 & \text{if } sub_{ik} = a_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

One can consider that $m^{A_i}(.|do(a_{ij}))$ is provided by a source of information and this latest expects that it will be a certain *bba*. Since the occurrence of the intervention is uncertain, the *bba* defined by applying Equation 4.20 is not appropriate. Accordingly, this source is seen as not fully reliable. In fact, even if the intervention succeeds to put its target into one specific value, its occurrence remains uncertain. A Bayesian *bba* expressing the actual values concerning the occurrence of the intervention has been computed with BetP as explained in the last step. It will be used to evaluate the reliability of the source.

When considering the case of an intervention forcing the variable A_i to take the value a_{ij}, the occurrence of the intervention in the form of other states does not matter. What it was

predicted by the source is an intervention certainly occurring at the state a_{ij}, $m^{DO}(do(a_{ij})) = 1$, whereas the actual belief about the occurrence of the intervention succeeding to put the variable A_i into the state a_{ij} is defined as $m(do(a_{ij})) = \alpha \in [0, 1]$. Since the degree of confidence in the reliability of a source can depend on the true value of the variable of interest, the difference between what is was predicted and the actual value is considered as its discounting factor defined as $1 - \alpha$.

As a consequence, the conditional distribution given the *DO* node is discounted by taking into account the reliability of each source, namely $\alpha_{do(a_{ij})}$. This information, will transform the conditional given the *DO* node from a certain *bba* into a weaker, less informative one. Hence, the new conditional *bba* of the target variable given the *DO* node becomes:

$$m^{A_i,\alpha_{do(a_{ij})}}(sub_{ik}|do(a_{ij})) = \begin{cases} 1-\alpha & \text{if } sub_{ik} = \{a_{ij}\} \\ \alpha & \text{if } sub_{ik} = \Theta_{A_i} \end{cases} \tag{4.21}$$

Proposition 4.9. *Standard interventions are a particular case of uncertain interventions when the source is fully reliable, i.e.,* $\alpha = 0$.

$$m^{A_i,\alpha_{do(a_{ij})}=0}(sub_{ik}|do(a_{ij})) = \begin{cases} 1 & \textit{if } sub_{ik} = \{a_{ij}\} \\ 0 & \textit{otherwise} \end{cases} \tag{4.22}$$

Example 4.10. *(continued)*

Graphically, an extra node DO representing the intervention on the variable C is added as its new parent. The values of this node are computed in the example. Each conditional distribution for the target variable C given an instance of the DO node is seen as provided by a distinct source of information. These sources affirm that performing an intervention will lead to a known change in the state of the manipulated variable. The conditional distributions as presented by the sources are presented in Table 4.7.

Table 4.7: Certain *bba*: $m^C(.|do(x))$

	$\{do(c_1)\}$	$\{do(c_2)\}$	$\{do(nothing)\}$
$\{c_1\}$	1	0	0
$\{c_2\}$	0	1	0
Θ_C	0	0	1

Since the intervention achievement is uncertain, conditional local distributions presented in Table 4.7 are not appropriate. In fact, even when the intervention occurs with a degree of belief and succeeds to put its target into one specific value, one should take into consideration the cases where it fails to take place. Therefore, certain conditional local distributions will be discounted according to the reliability of each source. The degree of confidence in the reliability of a source is computed according the true value of the variable of interest, i.e. the DO bba. Hence, discount rates are denoted by $1 - \alpha_{do(x)}$. *They are defined as 1-* $\alpha_{do(c_1)} = 0.245$, *1-* $\alpha_{do(c_2)} = 0.01$ *and 1-* $\alpha_{do(nothing)} = 0.745$.
The new discounted conditional bba given the DO node is presented in Table 4.8.

Table 4.8: Discounted *bba*: $m^{C,\alpha_{do(x)}}(.|do(x))$

	$\{do(c_1)\}$	$\{do(c_2)\}$	$\{do(nothing)\}$
$\{c_1\}$	1*0.245=0.245	0*0.01=0	0*0.745=0
$\{c_2\}$	0*0.245=0	1*0.01=0.01	0*0.745=0
Θ_C	0*0.245+0.755 =0.755	0*0.01+0.99=0.99	1*0.745+0.255=1

4- Defining conditionals given an uncertain intervention

The impact of the uncertain intervention on the target variable will not only depend from the intervention but also from the initial causes of the variable. The Dempster rule of combination is used to aggregate the conditional distribution given the initial causes with the discounted conditional given the *DO* parent. We use $m^{A_i}(a_j|Pa(A_i))$ to represent the conditional mass function induced on the space Θ_{A_i} given $Pa(A_i) \subseteq \Theta_{PA(A_i)}$, and $m^{A_i,\alpha_{do(x)}}(a_k|do(x))$ to represent the discounted conditional mass function induced on the space Θ_{A_i} given the intervention $do(x)$. The resulting *bba* of the target variable $m^{A_i}(a_i|Pa(A_i), do(x))$ is computed as follows:

$$m^{A_i}(a_i|Pa(A_i), do(x)) = \sum_{a_j \cap a_k = a_i} m^{A_i}(a_j|Pa(A_i)) \cdot m^{A_i,\alpha_{do(x)}}(a_k|do(x)) \quad (4.23)$$

Example 4.11. *(continued)*

The mass distributions given the initial causes and the DO node will be aggregated to give the conditional bba $m^C(.|s_i, do(x))$*. This latter represents the effect of an uncertain intervention on the variable C. For instance,* $m^C(.|s_1, do(c_1))$ *is obtained by computing* $m^C(.|s_1) \oplus m^{C,\alpha_{do(c_1)}}(.|do(c_1))$*. Results are presented in Table 4.9.*

Unlike the case of standard interventions, $m^C(c_1|s_1, do(c_1)) \neq 1$*. However, the action of the friend has raised the beliefs about the sweetness of the coffee. A small increase from 0.8 to 0.845 is explained by the fact that it is more likely that the used ingredient is salt. In the same way,* $m^C(c_2|s_2, do(c_1))$ *has decreased from 0.7 to 0.638.*

Table 4.9: Conditional *bba*: $m^C(.|s_i, do(c_1))$

	$\{(s_1, do(c_1))\}$	$\{(s_2, do(c_1))\}$
$\{c_1\}$	0.8450	0.180
$\{c_2\}$	0.0775	0.638
Θ_C	0.0775	0.182

Interventions not occurring

The approach proposed for interventions uncertainly happening is also valid in the case when no interventions are performed. It amounts to compute the effect of observations on the system. This is represented by setting the variable *DO* with certainty to the value $do(nothing)$.

The situation of non-intervention encompasses:

- not acting on the target variable and observing the spontaneous behavior of the system,
- failing to act on the target variable and therefore the intervention will not occur.

Formally, in this case:

$$\forall \theta_i, match(\theta_i) = \{nothing\} \tag{4.24}$$

From Equations 4.19 and 4.24, the *bba* of the *DO* node is defined by:

$$m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = \{nothing\} \\ 0 & \text{otherwise} \end{cases} \tag{4.25}$$

In this case, the state of the target variable will not depend on the intervention (i.e., from the *DO* node). The conditional *bba* given the *DO* node is not informative. It is represented with the vacuous *bba* defined as:

$$m^{A_i}(sub_{ik}|do(nothing)) = \begin{cases} 1 & \text{if } sub_{ik} = \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases} \tag{4.26}$$

The "Non-intervention" occurs certainly. Therefore, the source is fully reliable and the discounting factor is equal to zero. Hence, our approach well handles the particular case of standard interventions.

Proposition 4.10. *The beliefs provided about the non-occurrence of an intervention are accepted without any modification. They are defined like standard interventions.*

$$m^{A_i,\alpha_{do(nothing)}}(.|do(nothing)) = m^{A_i}(.|do(nothing)) \tag{4.27}$$

The conditional *bba*s defined in the context of the *DO* node and of the initial causes are computed by combining each conditional defined per single parent as follows:

$$\begin{aligned} m^{A_i}(.|Pa(A_i), do(nothing)) &= m^{A_i}(.|do(nothing)) \oplus m^{A_i}(.|Pa(A_i)) \\ &= m^{A_i}(.|Pa(A_i)) \end{aligned} \tag{4.28}$$

Proposition 4.11. *An augmented belief function causal graph where the DO node is set to the value nothing encodes the same joint distribution than the initial causal belief network.*

$$m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^{V}_{\mathcal{G}} \tag{4.29}$$

where $V' = V \cup DO$.

4.5 Handling interventions with imperfect consequences

In the last section, we have dealt with interventions occurring in an imperfect way. When happening, even with a belief $m(\{do(a_{ij})\})$, they succeed to put the target variable into exactly one specific state which is not often feasible. Therefore, our proposed approach in this section is to handle interventions with imperfect consequences, i.e., occurring and failing to put their target into a specific value (Boukhris, Elouedi, & Benferhat, 2012b). Graphically, they are represented on augmented belief networks.

4.5.1 Standard interventions with imperfect consequences

We propose to specify a new *bba* on the target variable representing the consequences of the intervention. Let us denote by $\mathcal{F}_{A_i}$, the set of the focal elements representing the uncertain consequences of the intervention where a *bbm* β_j is allocated to each focal element. The conditional *bba* of the target variable given a standard intervention on the variable A_i attempting to force it to take the value a_{ij} is defined as follows:

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} \beta_j & \text{if } sub_{ik} \in \mathcal{F}_{A_i}, \beta_j \in]0,1]; \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

Example 4.12. *(continued)*

Let us continue with the example of saccharine. Even if the substance is a kind of sugar, adding it will obviously affect the sweetness of the coffee but without certainty. The conditional bba $m^C(.|do(c_1))$ reflecting the impact of this action is presented in Table 4.10.

Table 4.10: Local *bba* upon a standard intervention: $m^C(.|do(c_1))$

	$\{\text{do}(c_1)\}$
$\{c_1\}$	0.8
$\{c_2\}$	0.05
Θ_C	0.15

4.5.2 Imperfect interventions with imperfect consequences

As argued in the last section, when handling uncertain interventions succeeding to set their target into a specific value a_{ij}, the conditional *bba*s given instances of the *DO* node will be discounted according to the actual occurrence of the intervention (see Equation 4.21). Here we consider the case of interventions can take several possible states. Therefore, the resulting *bba* is defined as a mixture of Equation 4.21 and 4.30 as follows:

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} (1-\alpha)\cdot\beta_j & \text{if } sub_{ik} \in \mathcal{F}_{A_i} \\ \alpha + (1-\alpha)\cdot\beta_j & \text{if } sub_{ik} = \Theta_{A_i}. \end{cases} \tag{4.31}$$

Proposition 4.12. *Uncertain interventions with a certain consequence are a particular case of uncertain ones with uncertain consequences when the parameter β_j is set to one.*

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} 1-\alpha & \textit{if } sub_{ik} = \{a_{ij}\} \\ \alpha & \textit{if } sub_{ik} = \Theta_{A_i}. \end{cases}$$

Example 4.13. *(continued)*

As stated in the first example, since you are in a restaurant it is more likely that what your friend has putted into your coffee is salt. We are focusing in the occurrence of the intervention as attempting to set its target into the value sweet, which means that the powder is sugar. However as mentioned in the last example, some kinds of sugar (e.g., lactose, saccharine) are either not very soluble or may have a bitter or metallic unpleasant aftertaste. Adding them may have uncertain consequences. Since the bbm that the added substance is sugar is represented with $m(\{do(c_1)\})$= 0.245. To represent this case of uncertain intervention with

uncertain consequences, the conditional bba given the DO node presented in Table 4.10 will be discounted. The resulting bba is presented in Table 4.11:

Table 4.11: Local *bba* upon an uncertain intervention: $m^{C,\alpha_{do(c_1)}}(.|do(c_1))$

	$\{\text{do}(c_1)\}$
$\{c_1\}$	$0.8 \cdot 0.245 = 0.2$
$\{c_2\}$	$0.05 \cdot 0.245 = 0.01$
Θ_C	$0.15 \cdot 0.245 + 0.755 = 0.79$

As for uncertain interventions with certain consequences, the conditional distribution given the *DO* parent will be combined with the discounted conditional distribution given the initial causes using Dempster's rule of combination to obtain the actual distribution of the target variable upon acting on it.

4.6 Conclusion

This chapter provided a graphical model to deal with interventions under a normalized belief framework, namely the causal belief network *(CBNC)*. This network provides an appropriate model to represent imperfect causal knowledge in particular ignorance situations. We have first presented an approach to define belief causation and represent the effect of interventions. We have shown that in order to correctly represent causal relations and reason in a causal way, the structure of the network has to be modified and the conditioning on observation should be distinguished from a conditioning on an external action. A generalization of the "do" operator under the belief function framework was therefore proposed, mutilated and augmented belief graphs were shown to be equivalent methods even if the joint distribution is not defined as for probability distribution. We have also proposed a new representation of external actions, namely, imperfect interventions that may have imperfect consequences. We have shown that they have a natural encoding under the belief function framework.

In the next chapter we will see how an agent may use this belief function causal network to ascribe causality between a sequence of reported events.

Chapter 5

Causality ascription

5.1 Introduction

Understanding what happens in a sequence of time-stamped reported events is important in several applications (e.g., surveillance problems, the modeling of how people may perceive the information with which they are faced). This task involves the identification of relevant patterns involving abnormal events, and requires the use of background knowledge (Chassy et al., 2011).

The notion of causality plays a central role in such problems. The intelligent artifact can ascribe as causal the link between reported events in a sequence and therefore can predict what event is likely to take place. Hence the need to have models of causality ascriptions. As explained in Chapter 2, the task of ascribing causality is not the one of diagnosis. Besides, it is different from prediction problems (Benferhat et al., 2008). It consists in inferring an unknown causal relation from two known events and some background non-causal knowledge.

A prerequisite for a relevant definition of causality ascription is a language for describing the agent's background knowledge. This knowledge should tolerate exceptional situations (Bonnefon et al., 2008). It seems clear that the ascription of causality depends on the choice of this language.

The non-monotonic approach, introduced in Chapter 2, responds to this condition. It allows to identify causal relationships among observed events. However, it is appropriate only for qualitative knowledge and as a drawback of this approach, the representation of events is restrained to binary variables. An intelligent artifact should also be able to reason from uncertain quantitative information and a set of observations or interventions occurring in his environment in order to ascribe causality.

In this chapter, we are interested in ascribing causality when the agent's background knowledge is expressed with the belief function formalism. This latter is well known for its expressive power and consequently, it is appropriate to deal with the poor knowledge of the world that agents usually possess. It well represents cases of partial and total ignorance and overcomes the limitation of qualitative models (Bonnefon et al., 2006).

In the context of observations, any representation of mass assignments is suitable. In the context of interventions, having a graphical structure or the use of structural equations (Halpern & Pearl, 2005) are needed. Since incomplete or insufficient information may let the construction of structural equation impossible, we will use the causal belief network presented in Chapter 4 to describe the background knowledge of the agent in the presence of interventional data.

We propose a model that an intelligent agent will use to ascribe causality from a sequence of observations or interventions occurring in his environment under the belief function framework. In fact, he will look for the causes of an exceptional event that have changed the normal course of things to an abnormal situation. To model such changes, we use the concepts of acceptance and rejection instead of changes in uncertainty distributions. Several levels of acceptance and rejection are introduced. Different alternative definitions of causality ascription, based on the specification of the strength of the cause, can be provided according to the different levels of acceptance and rejection.

This chapter is consecrated to the ascription of causality under the belief function framework. Since our work is based on the concepts of acceptance and rejection, we introduce their definition in Section 5.2. Section 5.3 details the causality ascription model in presence of observations and Section 5.4 shows how to ascribe causality in presence of interventions. Facilitation and justification are two concepts very related to causality. Thus, events related in a causal way should be well distinguished from those when facilitation or justification are involved. We explain how to identify them in presence of observations in Section 5.5 and in presence of interventions in Section 5.6. Proofs of this chapter are given in Appendix B.

5.2 The role of acceptance, rejection in causality ascription

5.2.1 (Ab)normality and ascribing causality

The notion of abnormality as discussed by philosophers of law (Hart & Honoré, 1985) and experimentally checked by psychologist (Hilton & Slugoski, 1986), is privileged when providing causal explanations. This concept is central in our causality ascription model.

Notice that we are not looking for actual causes. As explained in the following, we want to predict unknown causal relations that an agent will ascribe based on three components namely the knowledge he holds, a set of events reported to him and an *abnormal event*:

- his non-causal background knowledge about the natural course of the world expressed here in the belief function formalism which is a general and appropriate framework to model the world imperfection. Remember that masses take numerical values in [0,1].
- a set of observations or interventions occurring in his environment denoted by $\{f_1, \ldots, f_n\}$ such that $f_i \subseteq \Theta$ where Θ is the cartesian product of all n-ary variable domains;
- an event that contradicts its judgment about the normal course of things: an abnormal event $e_i \subseteq \Theta$. This event is involved in the causal process. In fact, it is an effect for which the agent is looking for its possible causes.

Let $\Theta_E = \{e_1, e_2, \ldots, e_n\}$ be the set of events defined as a partition of Θ such that events $e_i \in \Theta_E$ are *exhaustive and mutually exclusive*:

$$\begin{cases} Exhaustibility : e_1 \cup e_2 \cup \ldots \cup e_n = \Theta, \\ Exclusivity : \forall i, j \;\; e_i \cap e_j = \emptyset. \end{cases}$$

The complement of e_i w.r.t. Θ_E, denoted by $\overline{e_i}$ is defined as:

$$\overline{e_i} = \bigcup_{e_j, e_j \neq e_i} e_j.$$

A very particular case concerns atomic events of the form $A_i = \{a_{ij}\}$. In this case $\Theta_E = \{[a_{i1}], \ldots, [a_{in}]\}$ where $[a_{ij}]$ is a set of all elements $\theta \in \Theta$ such that $\theta^{\downarrow A_i} = \{a_{ij}\}$.

Thus, ascribing causality under the belief function theory consists in determining among *temporally sequenced* events f_i the ones that, in a given *context*, causes the occurrence of the abnormal event e_i. When there is no ambiguity, all events including abnormal ones will be denoted by e_i.

In our model, we propose to use the concepts of acceptance, rejection and ignorance to assert that events are bounded by a causal relation (Boukhris, Benferhat, & Elouedi, 2011a) such that an event is considered as:

- normal if it is an accepted event;
- abnormal if it is a rejected event.

Example 5.1. *Assume that your friend has a sore throat which is exceptional for him. Actually in the normal course of things, it is a rejected event. You want to ascribe the causes of this ill. It was reported to you that the day before he had taken a cold shower, had gone out afterwards and that he had talked too long too loud. They are the potential causes of suffering from a sore throat and we need to discriminate between them.*

An event may have different possible states namely accepted, rejected or ignored in a given context. The following subsections provide detailed explanations of these concepts. Indeed, we propose more different levels of acceptance and rejection (Boukhris, Elouedi, & Benferhat, 2012a). These notions are used to describe the state of potential causes as well as the rejected effect e_i.

5.2.2 Acceptance

An event is considered as accepted if it is likely enough to be considered as it holds. We propose to distinguish between three levels of acceptance under the belief function framework based on the plausibility of the event e_i and its relation with the plausibility of other events constituting its complement. An event e_i can be either very strongly, strongly or weakly accepted in a given context.

Very strong acceptance

In a given context, an event e_i is very strongly accepted if the confidence in this event is strictly greater than the confidence in its complement:

$$pl(e_i) > pl(\overline{e_i})$$

Strong acceptance

An event e_i is strongly accepted if it is the only one that has the highest plausibility (the best confidence). Hence, it is *exclusively* in $Argmax_{e_j \in \Theta_E} pl(e_j)$:

- $pl(e_i) = Argmax_{e_j \in \Theta_E}(pl(e_j))$;
- $\forall e_k \neq e_i \in \Theta_E$, $\text{pl}(e_k) \neq Argmax_{e_j \in \Theta_E}(\text{pl}(e_j))$

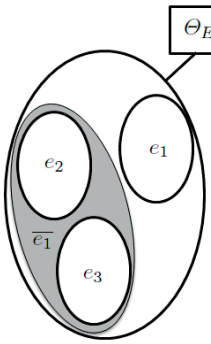
Weak acceptance

An event e_i is considered as weakly accepted, if it has the highest plausibility but there are also *other events* that share its confidence (i.e., in $Argmax_{e_j \in \Theta_E}$ $\text{pl}(e_j)$), formally:

- $pl(e_i) = Argmax_{e_j \in \Theta_E}(pl(e_j))$;
- $\exists e_k \neq e_i \in \Theta_E$, $\text{pl}(e_k) \neq Argmax_{e_j \in \Theta_E}(\text{pl}(e_j))$

Example 5.2. *An agent would like to judge if not having a sore throat is an accepted event.*

The frame of discernment associated to the variable having sore throat is denoted by Θ_S $= \{s_1$=*usually,* s_2=*yes,* s_3=*no*$\}$. *Beliefs are defined on* 2^{Θ_S}. *Let us suppose that events are* $e_1=\{s_1\}$, $e_2=\{s_2\}$ *and* $e_3=\{s_3\}$. *We consider beliefs presented in the three rows of Figure 5.1:*



$\text{pl}(e_1)$	$\text{pl}(e_2)$	$\text{pl}(e_3)$	$\text{pl}(\overline{e_1})$	States of e_1
1	0.64	0.24	**0.73**	very strong accept
0.4	0.3	0.3	0.6	strong accept
0.5	**0.5**	0.1	0.6	weak accept

Figure 5.1: Forms of acceptance

1. $pl(e_1) > pl(\overline{e_1})$*:the plausibility of the event* e_1 *is greater than the the plausibility of its complement (i.e.,* $\{s_2, s_3\}$*). In this case* e_1 *is considered as a* ***very strongly accepted*** *event.*

2. $pl(e_1) > pl(e_2)$ *and* $pl(e_1) > pl(e_3)$*: the event* e_1 *has the maximal plausibility value namely* $pl(e_1)$=$Argmax(pl(e_i))$. *Since only* $pl(e_1) = Argmax(pl(e_i))$, *then* e_1 *is perceived as a* ***strongly accepted*** *event.*

3. $pl(e_1) = pl(e_2)$ *and* $pl(e_1) > pl(e_3)$*: the event* e_1 *has the maximal plausibility value namely* $pl(e_1) = Argmax(pl(e_i))$. *The agent has compared the plausibility of the remaining events with the value of* $Argmax(pl(e_i))$ *and has found that* $pl(e_2)$ *is also equal to* $Argmax(pl(e_i))$. *Therefore, he considers* e_1 *as a* ***weakly accepted*** *event.*

5.2.3 Ignorance

When an agent has the same confidence in any event e_i, then the validity of e_i is ignored. Its occurrence, is therefore as plausible as the occurrence of any other alternative. Formally, e_i is ignored if:

$$\forall e_i \in \Theta_E, pl(e_i) = pl(\overline{e_i})$$

5.2.4 Rejection

By symmetry to acceptance, different levels of rejection can be defined:

Very strong rejection

An event e_i is very strongly rejected if the plausibility of this event is strictly less than the plausibility of its complement:

$$pl(e_i) < pl(\overline{e_i})$$

Strong rejection

An event e_i is strongly rejected if:

- $pl(e_i) = Argmin_{e_j \in \Theta_E}(pl(e_j))$;
- $\forall e_k \neq e_i \in \Theta_E$, $pl(e_k) \neq Argmin_{e_j \in \Theta_E}(pl(e_j))$

Weak rejection

An event e_i is defined as weakly rejected if:

- $pl(e_i) = Argmin_{e_j \in \Theta_E}(pl(e_j))$;
- $\exists e_k \neq e_i \in \Theta_E$, $pl(e_k) \neq Argmin_{e_j \in \Theta_E}(pl(e_j))$

In the following, we provide a characterization of different concepts of acceptance and rejection introduced above:

Definition 5.1. *An event $e_i \in \Theta_E$ is perceived as:*

- *very strongly accepted if $pl(e_i) > pl(\overline{e_i})$*
- *strongly accepted if:*
  - *$pl(e_i) = Argmax_{e_j \in \Theta_E}(pl(e_j))$;*
  - *$\forall e_k \neq e_i \in \Theta_E$, where $pl(e_k) \neq Argmax_{e_j \in \Theta_E}(pl(e_j))$*
- *weakly accepted if:*
  - *$pl(e_i) = Argmax_{e_j \in \Theta_E}(pl(e_j))$;*
  - *$\exists e_k \neq e_i \in \Theta_E$, $pl(e_k) \neq Argmax_{e_j \in \Theta_E}(pl(e_j))$*
- *very strongly rejected if $pl(e_i) < pl(\overline{e_i})$*

- *strongly rejected if:*
  - $pl(e_i) = Argmin_{e_j \in \Theta_E}(pl(e_j))$;
  - $\forall e_k \neq e_i \in \Theta_E$, $pl(e_k) \neq Argmax_{e_j \in \Theta_E}(pl(e_j))$
- *weakly rejected if:*
  - $pl(e_i) = Argmin_{e_j \in \Theta_E}(pl(e_j))$;
  - $\exists e_k \neq e_i \in \Theta_E$, $pl(e_k) \neq Argmin_{e_j \in \Theta_E}(pl(e_j))$

As stated in the following proposition, if we restrict our definitions to the special case of binary partitions, i.e., $\Theta_E = \{e_i, \overline{e_i}\}$, all definitions of acceptance (resp. rejection) are equivalent.

Proposition 5.1. *If $\Theta_E = \{e_i, \overline{e_i}\}$ then:*

- *e_i is weakly accepted iff e_i is strongly accepted iff e_i is very strongly accepted iff $pl(e_i) > pl(\overline{e_i})$.*
- *e_i is weakly rejected iff e_i is strongly rejected iff e_i is very strongly rejected iff $pl(e_i) < pl(\overline{e_i})$.*

5.2.5 Conditional acceptance and conditional rejection

Concepts of acceptance, rejection can be extended in order to take into account a given context. The context of an event is the set of the circumstances and conditions surrounding it. Therefore, it is a set of events occurring in the normal course of things. The context, will be denoted by c. In this subsection, we introduce different forms of conditional acceptance and rejection.

Conditional acceptance

Conditional very strong acceptance After the occurrence of the event c, e_i becomes very strongly accepted if:

$$pl(e_i|c) > pl(\overline{e_i}|c)$$

Conditional strong acceptance After the occurrence of the event c, e_i becomes strongly accepted if:

- $pl(e_i|c) = Argmax_{e_k \in \Theta_E}(pl(e_k|c))$;
- $\forall e_j \neq e_i \in \Theta_E$, $pl(e_j|c) \neq Argmax_{e_k \in \Theta_E}(pl(e_k|c))$

Conditional weak acceptance After the occurrence of the event f_j, e_i becomes weakly accepted if:

- $pl(e_i|c) = Argmax_{e_k \in \Theta_E}(pl(e_k|c))$;
- $\exists e_j \neq e_i \in \Theta_E$, $pl(e_j|c) \neq Argmax_{e_k \in \Theta_E}(pl(e_k|c))$

Conditional rejection

Conditional very strong rejection After the occurrence of the event c, e_i becomes very strongly rejected if:

$$pl(e_i|c) < pl(\overline{e_i}|c)$$

Conditional strong rejection After the occurrence of the event c, e_i becomes strongly rejected if:

- $pl(e_i|c) = Argmin_{e_k \in \Theta_E}(pl(e_k|c))$;
- $\forall e_j \neq e_i \in \Theta_E$, $pl(e_j|c) \neq Argmin_{e_k \in \Theta_E}(pl(e_k|c))$

Conditional weak rejection After the occurrence of the event c, e_i becomes weakly rejected if:

- $pl(e_i|c) = Argmin_{e_k \in \Theta_E}(pl(e_k|c))$;
- $\exists e_j \neq e_i \in \Theta_E$, $pl(e_j|c) \neq Argmin_{e_k \in \Theta_E}(pl(e_k|c))$

5.3 Ascribing causality from a belief background knowledge

In this section, we propose a model to ascribe causality between events when the background knowledge is uncertain and expressed under the belief function framework.

5.3.1 Definition

Causes are necessary to the occurrence of their effects, a causal link defines a higher belief of effects when a cause takes place. Thus, if a cause does not arise then the plausibility of the occurrence of the effect will decrease. Note that a delay is required for the cause to make its effect happening. Hence at a given context, if the cause appears at time t, then its effects occur later at time $t + n$.

In order to ascribe the causes of the abnormal event e_i, the agent has to select events that promote its acceptance. The event e_i is rejected in the normal course of things in a given context c at time t. Events from a sequence of observations or interventions are the potential causes of e_i. Indeed, if the event e_i becomes accepted in the same context c after the occurrence of e_j at time $t + n$, then e_j is seen as a cause of e_i. Causal inference (propagating beliefs in the CBNC), allows to compute the plausibility of e_i in context c at time t, namely $pl^t(e_i|c)$ and the plausibility of e_i in context c after the occurrence of e_j at time $t+n$, namely $pl^{t+n}(e_i|e_j, c)$. Note that the difference between the level of rejection of the event e_i in the normal course of things and its acceptance after the occurrence of e_j determines the *strength of the causal link*.

5.3.2 Ascribing causes in presence of observational data

As mentioned above, the strength of the causal relation depends on the difference between the level of rejection of the effect which is an abnormal event denoted by e_i and its acceptance after the occurrence of the causes e_j. For instance as depicted in Figure 5.2, the difference

between very strong rejection and strong acceptance is equal to the difference between strong rejection and very strong acceptance and is lesser than the difference between very strong rejection and very strong acceptance.

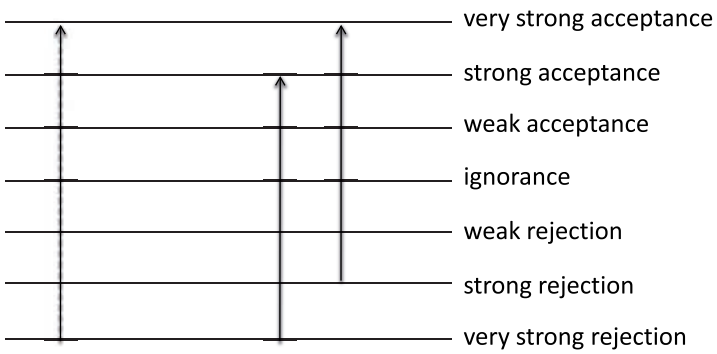


Figure 5.2: Difference between levels of rejection and acceptance

Consequently, we obtain five strengths that we called very strong, strong, weak, very weak and slight causes (see Table 5.1).

Table 5.1: Strengths of the causal link

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

Ascribing very strong causes

As shown in Table 5.2 very strong causes can be identified only from very strongly abnormal events.

Table 5.2: Ascribing very strong causes

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

-1- from a very strongly abnormal effect Two events are perceived as very strongly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+n$, after observing the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.2. *If an event e_i is very strongly rejected in a given context and* after *observing an event e_j it becomes very strongly accepted, then e_j is said to be a* very strong cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c); \tag{5.1}$$

$$pl^{t+n}(e_i|e_j,c) > pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.2}$$

Example 5.3. *Assume that an agent has in his disposal some information expressed with the belief function formalism: "generally, airlines do not delay their flights". Flights delay is represented with Θ_F=\{yes (y), no (n)\} where $pl^t(\{y\})$ =0.3, $pl^t(\{n\})$ =0.8 and $pl^t(\Theta_F)$ =1. We notice that $pl^t(\{y\})$ is an abnormal event that is very strongly rejected since $pl^t(\{y\})$ =0.3<$pl^t(\{n\})$ =0.8.*
From the observation: "The eruption of the Eyjafjöll volcano creates high quantity of ash clouds. Afterwards, all European countries delayed their flights", the agent should be able to identify if it exists a causal relation between ash clouds and flights delay. The quantity of ash clouds after the eruption of a volcano is represented with Θ_A=\{high (h)> $4mg/m^3$, $2mg/m^3$≤medium (m)≤$4mg/m^3$, low (l)< $2mg/m^3$\}.
In context of high ash clouds, flights delay becomes very strongly accepted $pl^{t+n}(\{y\}|\{h\})$ =0.9 > $pl^{t+n}(\{n\}|\{h\})$ =0.2. Hence, the high quantity of ash clouds can be seen as a very strong cause of flights delay.

Ascribing strong causes

Strong causes are identified in two cases according to the degree of abnormality of the effect. In fact, as shown in Table 5.3 very strong causes can be identified from very strongly abnormal events or from strongly abnormal ones.

Table 5.3: Ascribing strong causes

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

-1- from a very strongly abnormal effect Two events are perceived as strongly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+n$, after observing the other one, he changes his beliefs and strongly accepts it.

Proposition 5.3. *If an event e_i is very strongly rejected in a given context and* after *observing an event e_j it becomes strongly accepted, then e_j is said to be a* strong cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c); \tag{5.3}$$

$\exists e \neq e_i$,

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.4}$$

-2- from a strongly abnormal effect Two events are perceived as strongly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+n$, after observing the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.4. *If an event e_i is strongly rejected in a given context and* after *observing an event e_j it becomes very strongly accepted, then e_j is said to be a* strong cause *of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \tag{5.5}$$

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.6}$$

Ascribing weak causes

Weak causes are identified in three cases according to the abnormality of the effect as shown in Table 5.4.

Table 5.4: Ascribing weak causes

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

-1- from a very strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+1$, after observing the other one, he changes his beliefs and weakly accepts it.

Proposition 5.5. *If an event e_i is very strongly rejected in a given context and* after *observing an event e_j it becomes weakly accepted, then e_j is said to be a* weak cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c) \tag{5.7}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.8}$$
$$where \;\; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

-2- from a strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+1$, after observing the other one, he changes his beliefs and strongly accepts it.

Proposition 5.6. *If an event e_i is strongly rejected in a given context and* after *observing an event e_j it becomes strongly accepted, then e_j is said to be a* weak cause *of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \tag{5.9}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.10}$$
$$where \;\; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

-3- from a weakly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+1$, after observing the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.7. *If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes very strongly accepted, then e_j is said to be* a weak cause *of e_i namely,*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.11}$$

$$pl^{t+n}(e_i|e_j,c) > pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.12}$$

Ascribing very weak causes

As shown in Table 5.5, very weak causes are identified in two cases according to the abnormality of the effect.

Table 5.5: Ascribing very weak causes

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

-1- from a strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+n$, after observing the other one, he changes his beliefs and weakly accepts it.

Proposition 5.8. *If an event e_i is strongly rejected in a given context and after observing an event e_j it becomes weakly accepted, then e_j is said to be a very weak cause of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \tag{5.13}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.14}$$
$$where \;\; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

-2- from a weakly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+n$, after observing the other one, he changes his beliefs and strongly accepts it.

Proposition 5.9. *If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes strongly accepted, then e_j is said to be a very weak cause of e_i, namely*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.15}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.16}$$
$$where \;\; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

Ascribing slight causes

Slight causes are derived from only weakly abnormal events as depicted in Table 5.6.

Table 5.6: Ascribing slight causes

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	very strongly accepted	very strong cause
	strongly accepted	strong cause
	weakly accepted	weak cause
strongly rejected	very strongly accepted	strong cause
	strongly accepted	weak cause
	weakly accepted	very weak cause
weakly rejected	very strongly accepted	weak cause
	strongly accepted	very weak cause
	weakly accepted	slight cause

-1- from a weakly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+n$, after observing the other one, he changes his beliefs and weakly accepts it.

Proposition 5.10. *If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes weakly accepted, then e_j is said to be a slight cause of e_i, namely*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.17}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c) \tag{5.18}$$
$$where \;\; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

5.3.3 Ascribing causes in presence of interventional data

Observational data provide some information about the statistical relations among events. It means that spurious correlation can be identified as events causally connected. In fact, two events may be wrongly inferred as causally related, due to either the coincidence of their occurrence or the presence of a common cause which is a hidden event.

Example 5.4. *An agent learns that someone took up drugs, that he has dilated pupils. He notices that this person's heart rate has increased. The agent believes that generally, it is abnormal to be a drug-consumer, to have dilated pupils, and to have an accelerated heart rate. If we consider an abnormal event as a very strongly rejected one we have:*
- $pl^t(\{\overline{Drugs}\}) > pl^t(\{Drugs\})$;
- $pl^t(\{\overline{Dilated}\}) > pl^t(\{Dilated\})$;
- $pl^t(\{\overline{Accelerated}\}) > pl^t(\{Accelerated\})$.

From the observation: "a person who has dilated pupils, has also an accelerated heart rate", the agent will conclude that when pupils are dilated, it very strongly causes an increase in heart rate which is not actually the case.

- $pl^{t+n}(Accelerated|Dilated) > pl^{t+n}(\overline{Accelerated}|Dilated)$.
Tropicamide shortly acts on the dilation of the pupil. When it is applied as eyes drops, it forces the eyes to be dilated ($do(Dilated)$). After using this substance, the agent notes that his action has no effect on the speed of the heartbeat. Accordingly, he concludes that there is not a causal relation between these two events.

On the other hand, the agent believes that it is normal for a drug-consumer to have dilated pupils and to have an accelerated heart rate:
- $pl^{t+n}(Dilated|Drugs) > pl^{t+n}(\overline{Dilated}|Drugs)$;
- $pl^{t+n}(Accelerated|Drugs) > pl^{t+n}(\overline{Accelerated}|Drugs)$.
After forcing someone to take drugs ($do(Drugs)$), he observes that his pupils are dilated and the speed of his heartbeat is altered. Therefore, he concludes that the hidden event, namely taking drugs, is their common cause.

Accordingly, ascribing the cause of an event will be much better and easier and if it is based on data collected via active interventions (Boukhris, Benferhat, & Elouedi, 2012a) rather than passive observations. In the context of observations any representation of the background knowledge is suitable whereas in the context of interventions the graphical structure is needed. Interventions will be represented on this causal structure by the mean of the "do" operator. As for observational data, causes are defined given their strength (Boukhris, Benferhat, & Elouedi, 2012b).

Ascribing very strong causes

-1- from a very strongly abnormal effect Two events are perceived as very strongly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+n$, after acting the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.11. *If an event e_i is very strongly rejected in a given context and* after *acting on an event e_j it becomes very strongly accepted, then e_j is said to be a* very strong cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c); \tag{5.19}$$

$$pl^{t+n}(e_i|do(e_j), c) > pl^{t+n}(\overline{e_i}|do(e_j), c) \tag{5.20}$$

Ascribing strong causes

Strong causes are identified in two cases according to the degree of abnormality of the effect.

-1- from a very strongly abnormal effect Two events are perceived as strongly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+n$, after acting the other one, he changes his beliefs and strongly accepts it.

Proposition 5.12. *If an event e_i is very strongly rejected in a given context and* after *acting an event e_j it becomes strongly accepted, then e_j is said to be a* strong cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c); \tag{5.21}$$

$$\exists e \neq e_i, pl^{t+n}(e|do(e_j), c) < pl^{t+n}(e_i|do(e_j), c) \leq pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.22)$$
$$where\ pl^{t+n}(e_i|do(e_j), c) = Argmax(pl^{t+n}(e_k|do(e_j), c))$$

-2- from a strongly abnormal effect Two events are perceived as strongly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+n$, after acting the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.13. *If an event e_i is strongly rejected in a given context and* after *acting an event e_j it becomes very strongly accepted, then e_j is said to be a* strong cause *of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \quad (5.23)$$
$$pl^{t+n}(e_i|do(e_j), c) > pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.24)$$

Ascribing weak causes

Weak causes are identified in three cases according to the abnormality of the effect.

-1- from a very strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is very strongly rejected and at time $t+1$, after acting the other one, he changes his beliefs and weakly accepts it.

Proposition 5.14. *If an event e_i is very strongly rejected in a given context and* after *acting an event e_j it becomes weakly accepted, then e_j is said to be a* weak cause *of e_i, namely*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c) \quad (5.25)$$

$$\exists e \neq e_i, pl^{t+n}(e|do(e_j), c) \leq pl^{t+n}(e_i|do(e_j), c) \leq pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.26)$$
$$where\ \ pl^{t+n}(e_i|do(e_j), c) = Argmax(pl^{t+n}(e_k|do(e_j), c))$$

-2- from a strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+1$, after acting the other one, he changes his beliefs and strongly accepts it.

Proposition 5.15. *If an event e_i is strongly rejected in a given context and* after *acting an event e_j it becomes strongly accepted, then e_j is said to be a* weak cause *of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \quad (5.27)$$

$$\exists e \neq e_i, pl^{t+n}(e|do(e_j), c) < pl^{t+n}(e_i|do(e_j), c) \leq pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.28)$$
$$where\ \ pl^{t+n}(e_i|do(e_j), c) = Argmax(pl^{t+n}(e_k|do(e_j), c))$$

-3- from a weakly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+1$, after acting the other one, he changes his beliefs and very strongly accepts it.

Proposition 5.16. *If an event e_i is weakly rejected in a given context and after acting an event e_j it becomes very strongly accepted, then e_j is said to be* a weak cause *of e_i namely,*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \quad (5.29)$$
$$pl^{t+n}(e_i|do(e_j), c) > pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.30)$$

Ascribing very weak causes

Very weak causes are identified in two cases according to the abnormality of the effect.

-1- from a strongly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is strongly rejected and at time $t+n$, after acting the other one, he changes his beliefs and weakly accepts it.

Proposition 5.17. *If an event e_i is strongly rejected in a given context and after acting an event e_j it becomes weakly accepted, then e_j is said to be a very weak cause of e_i, namely*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \tag{5.31}$$

$$\begin{aligned} &\exists e \neq e_i, pl^{t+n}(e|do(e_j),c) \leq pl^{t+n}(e_i|do(e_j),c) \leq pl^{t+n}(\overline{e_i}|do(e_j),c) \\ &\qquad where \;\; pl^{t+n}(e_i|do(e_j),c) = Argmax(pl^{t+n}(e_k|do(e_j),c)) \end{aligned} \tag{5.32}$$

-2- from a weakly abnormal effect Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+n$, after acting the other one, he changes his beliefs and strongly accepts it.

Proposition 5.18. *If an event e_i is weakly rejected in a given context and after acting an event e_j it becomes strongly accepted, then e_j is said to be a very weak cause of e_i, namely* $\exists e \neq e_i$,

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.33}$$

$$\begin{aligned} &\exists e \neq e_i, pl^{t+n}(e|do(e_j),c) < pl^{t+n}(e_i|do(e_j),c) \leq pl^{t+n}(\overline{e_i}|do(e_j),c) \\ &\qquad where \;\; pl^{t+n}(e_i|do(e_j),c) = Argmax(pl^{t+n}(e_k|do(e_j),c)) \end{aligned} \tag{5.34}$$

Ascribing slight causes

Two events are perceived as weakly causally related, if the agent at time t, starts believing that one of them is weakly rejected and at time $t+n$, after acting the other one, he changes his beliefs and weakly accepts it.

Proposition 5.19. *If an event e_i is weakly rejected in a given context and after acting an event e_j it becomes weakly accepted, then e_j is said to be a slight cause of e_i, namely*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.35}$$

$$\begin{aligned} &\exists e \neq e_i, pl^{t+n}(e|do(e_j),c) \leq pl^{t+n}(e_i|do(e_j),c) \leq pl^{t+n}(\overline{e_i}|do(e_j),c) \\ &\qquad where \;\; pl^{t+n}(e_i|do(e_j),c) = Argmax(pl^{t+n}(e_k|do(e_j),c)) \end{aligned} \tag{5.36}$$

5.4 Attenuation and confirmation

Within the qualitative models (Bonnefon et al., 2006), if an event is held as accepted, then after the observation of a second event, it only may remain accepted or becomes rejected. Thus, attenuation and confirmation do not make sense. In qualitative possibilistic framework (Benferhat & Smaoui, 2008), the authors have shown that no distinction is made between weak independence and the case of confirmation and that the concept of attenuation cannot also be defined within that model. In the quantitative belief function framework acceptance, rejection can be confirmed or attenuated upon the occurrence of a new event.

5.4.1 Confirmation

Definition 5.2. *An event e_j is said to confirm another event e_i if the plausibility of observing e_i after observing e_j is greater than the plausibility of observing e_i alone.*

Proposition 5.20. *An event e_j is said to confirm another event e_i if:*

$$pl^t(e_i) \cdot pl^{t+n}(e_j) < pl^{t+n}(e_i, e_j) \tag{5.37}$$

5.4.2 Attenuation

Definition 5.3. *e_j is said to attenuate e_i if the plausibility of observing e_i after observing e_j is smaller than the plausibility of observing e_i alone.*

Proposition 5.21. *An event e_j is said to attenuate another event e_i if:*

$$pl^{t+n}(e_i, e_j) < pl^t(e_i) \cdot pl^{t+n}(e_j) \tag{5.38}$$

Example 5.5. *Suppose that an agent initial knowledge about the weather is:*
$pl^t(\{cold\}) = 0.8$, $pl^t(\{hot\}) = 0.1$ and $pl^t(\{cold, hot\}) = 1$. If later, he observes many people eating ice creams, I=$\{yes\}$, then his beliefs according to this new information are updated: $pl^{t+n}(\{(cold, yes)\}) = 0.4 < pl^{t+n}(\{cold\}) = 0.8$. He identifies eating ice creams as attenuating his belief about the cold weather.

5.5 Facilitation and justification in presence of observations

5.5.1 Facilitation

An agent deals with facilitation when he is cautious in his causal interpretation of the sequence of events. In fact, as for ascribing causality, to ascribe facilitation the agent starts not believing in the occurrence of an event and by observing another event, he changes his beliefs afterwards. However, this change consists not to believe in the event neither in its complement instead of accepting it as it is the case for causality. According the level of rejection of the event in the normal course of things, we got different levels of facilitations namely very strong, strong and weak. Table 5.7 shows these different strengths.

Table 5.7: Ascribing facilitations

State of the effect at time t	State of the effect at time t+n	Strength of the link
very strongly rejected	ignored	very strong facilitation
strongly rejected	ignored	strong facilitation
weakly rejected	ignored	weak facilitation

Very strong facilitation

Proposition 5.22. *If an event e_i is very strongly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to very strongly facilitate the occurrence of e_i. Namely,*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c). \tag{5.39}$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.40}$$

Example 5.6. *Let us continue with the Example 5.3 where flights delay is represented with Θ_F=$\{$yes (y), no (n)$\}$. The agent has at his disposal some background information such as $pl^t(\{y\})$ =0.3 < $pl^t(\{n\})$ =0.8 (a very strongly rejected event).*

Let us denote by $\Theta_S = \{s, \overline{s}\})$ the frame of discernment concerning airline strikes. After it was reported to him that some airlines strikes, the agent revises his beliefs and the event flights delay becomes ignored (i.e., $pl^{t+n}(y|s) = pl^{t+n}(n|s)$).

Accordingly, he will perceive the airline strikes, $\{s\}$, as very strongly facilitating the occurrence of flights delay,$\{y\}$. In fact, flights delay is unsurprising, but not expected to the agent.

Strong facilitation

Proposition 5.23. *If an event e_i is strongly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to strongly facilitate the occurrence of e_i. Namely,*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \tag{5.41}$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.42}$$

Weak facilitation

Proposition 5.24. *If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to weakly facilitate the occurrence of e_i. Namely,*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c) \tag{5.43}$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.44}$$

5.5.2 Justification

If an agent judges that the occurrence of an event gave reason to expect the occurrence of another event, we deal with justification. The first event caused the agent to start believing the second one. The agent should not be surprised of having it reported afterwards.

Justification is related to the notion of explanation. If the occurrence of an event e_i is ignored and after observing the occurrence of a second event e_j, it becomes expected, we deal then with justification. Accordingly, we obtain several levels of justification according to the levels of acceptance of the event. Table 5.8 shows these different strengths of justification.

Table 5.8: Ascribing justifications

State of the effect at time t	State of the effect at time t+n	Strength of the link
ignored	very strongly accepted	very strong justifciation
ignored	strongly accepted	strong justification
ignored	weakly accepted	weak justification

Very strong justification

Proposition 5.25. *If an event e_i is ignored in a given context and after observing an event e_j it becomes very strongly accepted then e_j is said to very strongly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \tag{5.45}$$

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.46}$$

Example 5.7. *Let Θ_W ={cold, hot, warm}. Assume that an agent ignores if the weather is cold pl^t({cold})= pl^t({warm, hot}) =0.5. After observing the event: "many persons are wearing coats", i.e., {coat}, the agent very strongly accepts that the weather is cold, i.e., {cold}. pl^{t+n}({(cold,coat)})=0.8 > pl^{t+n}({(hot,coat),(warm,coat)}) =0.1. In this case, he concludes that wearing coats very strongly justifies the cold weather.*

Strong justification

Proposition 5.26. *If an event e_i is ignored in a given context and after observing an event e_j it becomes strongly accepted then e_j is said to strongly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \tag{5.47}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c) \tag{5.48}$$
$$\text{where } \; pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

Weak justification

Proposition 5.27. *If an event e_i is ignored in a given context and after observing an event e_j it becomes weakly accepted then e_j is said to weakly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \tag{5.49}$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c) \quad (5.50)$$
$$where \;\; pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

5.6 Facilitation and justification in presence of interventions

In this section, we propose to ascribe facility and justification in presence of interventions. While with observational data any representation the background knowledge with mass functions is suitable, here it should be based on causal networks. Note that the details of the computation of the effect of an intervention on a belief function causal network was presented in Chapter 4.

5.6.1 Facilitation

Very strong facilitation

Proposition 5.28. *If an event e_i is very strongly rejected in a given context and after acting an event e_j it becomes ignored then e_j is said to very strongly facilitate the occurrence of e_i. Namely,*

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c). \quad (5.51)$$
$$pl^{t+n}(e_i|do(e_j), c) = pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.52)$$

Strong facilitation

Proposition 5.29. *If an event e_i is strongly rejected in a given context and after acting an event e_j it becomes ignored then e_j is said to strongly facilitate the occurrence of e_i. Namely,*

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c) \quad (5.53)$$
$$pl^{t+n}(e_i|do(e_j), c) = pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.54)$$

Weak facilitation

Proposition 5.30. *If an event e_i is weakly rejected in a given context and after acting an event e_j it becomes ignored then e_j is said to weakly facilitate the occurrence of e_i. Namely,*

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c) \quad (5.55)$$
$$pl^{t+n}(e_i|do(e_j), c) = pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.56)$$

5.6.2 Justification

Very strong justification

Proposition 5.31. *If an event e_i is ignored in a given context and after acting an event e_j it becomes very strongly accepted then e_j is said to very strongly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \quad (5.57)$$
$$pl^{t+n}(e_i|do(e_j), c) > pl^{t+n}(\overline{e_i}|do(e_j), c) \quad (5.58)$$

Strong justification

Proposition 5.32. *If an event e_i is ignored in a given context and after acting an event e_j it becomes strongly accepted then e_j is said to strongly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \tag{5.59}$$

$$\exists e \neq e_i, pl^{t+n}(e|do(e_j), c) < pl^{t+n}(e_i|do(e_j), c) \leq pl^{t+n}(\overline{e_i}|do(e_j), c) \tag{5.60}$$
$$where \;\; pl^{t+n}(e_i|do(e_j), c) = Argmax(pl^{t+n}(e_k|do(e_j), c))$$

Weak justification

Proposition 5.33. *If an event e_i is ignored in a given context and after acting an event e_j it becomes weakly accepted then e_j is said to weakly justify the occurrence of e_i. Namely,*

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c) \tag{5.61}$$

$$\exists e \neq e_i, pl^{t+n}(e|do(e_j), c) \leq pl^{t+n}(e_i|do(e_j), c) \leq pl^{t+n}(\overline{e_i}|do(e_j), c) \tag{5.62}$$
$$where \;\; pl^{t+n}(e_i|do(e_j), c) = Argmax(pl^{t+n}(e_k|do(e_j), c))$$

5.7 Conclusion

In this chapter, we have presented a model to ascribe causality between events. This model is a quantitative counterpart of the one presented in (Bonnefon et al., 2008) that is based on nonmonotonic consequence relations. In our model, we ascribe causality in presence of observations and also in presence of interventions. In fact, the use of the do operator avoids cases of spurious correlations. Besides, facility and justification were distinguished from causality in both cases of observational and interventional data. Attenuation and confirmation fully make sense under the belief function framework.

In the next chapter, we will implement our proposed causal belief function network and simulate the effect of an intervention using the belief graph mutilation and augmentation approaches. Methods regarding the belief causality ascription model will be developed. We will illustrate how we can ascribe the causes of an abnormal event from a sequence of reported events.

Chapter 6

Implementation and illustration

6.1 Introduction

In this chapter, we present the implementation of the causal belief network that we have proposed in this thesis. To this end, we develop programs in MATLAB V7.10.0 implementing our proposed methods to handle interventions in the causal belief network *(CBNC)* presented in Chapter 4. Experimentations show the effect of an intervention on the system by simulating the graph mutilation and the graph augmentation approaches.

Indeed, using the definitions proposed in Chapter 5, we illustrate the feasibility of our belief causality ascription model. Different results carried out from these simulations will be presented in order to show the usefulness of our proposed model. This chapter is organized into two parts as follows: Section 6.2 deals with the implementation of the causal belief network and experimentally demonstrates that upon an intervention the graph mutilation and graph augmentation approaches give the same results. Section 6.3 provides scenarios that are used to illustrate links between events. These latter are used to ascribe the possible causes of an abnormal event. In fact, an agent will be able to discriminate between potential causes. We report results of the different scenarios and provide an analysis of these results.

6.2 Causal belief networks

In this section, we present the causal belief network with conditional beliefs (CBNC) as a tool implemented in Matlab allowing the computation of the simultaneous effect of observations and interventions.

Note that this causal network is based on the definition of the associational network that we have proposed in Chapter 3 namely the (*BNC*). In the (CBNC), arcs follow the causal process. Besides, conditionals *bba*s are defined either per edge or for more than one parent nodes and the computation of the global joint distribution is based on the uniform ballooning and vacuous extensions concepts.

6.2.1 Graph creation

To create the *BNC* or *CBNC*, one should make:

1. *a qualitative specification:* it consists of defining the structure of the network. In this step, one should specify:

   - the number of nodes (see Figure 6.1).

Figure 6.1: Definition of the number of nodes

   - their labels and their cardinality (see Figure 6.2):

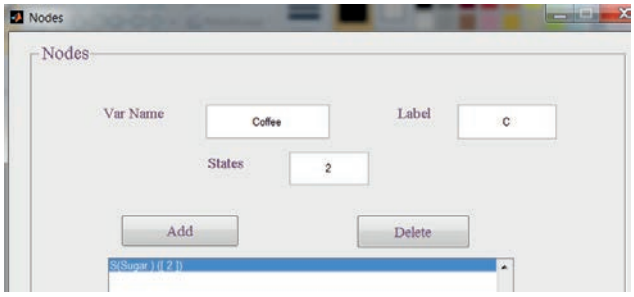


Figure 6.2: Definition of the labels and cardinality of the nodes

   - the set of their parents: when it is the *CBNC* that we want to define, this step corresponds to define the direct causes of a given node since arcs do not only represent dependence relations but also follow the causal process (see Figure 6.3):

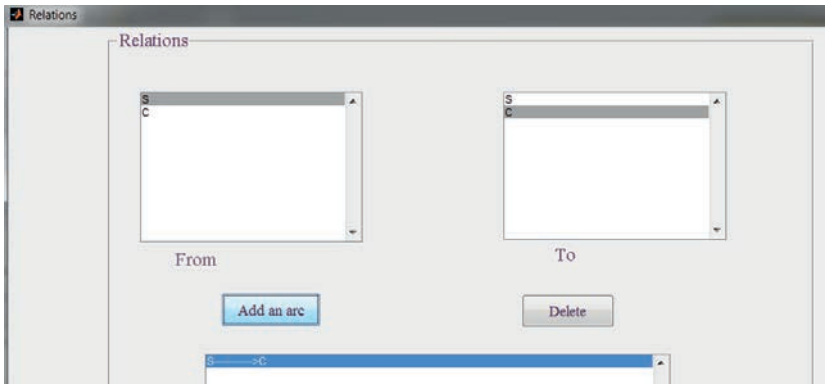


Figure 6.3: Definition of the causal links between nodes

2. *a quantitative specification*: once the structure of the causal network of Example 3.1 is defined, one should define local a priori mass distributions. Thus, a conditional distribution of each node in the context of either one or more parent nodes (see Figure 6.4 and Figure 6.5) have to be specified.

Figure 6.4: A priori local distribution m^S

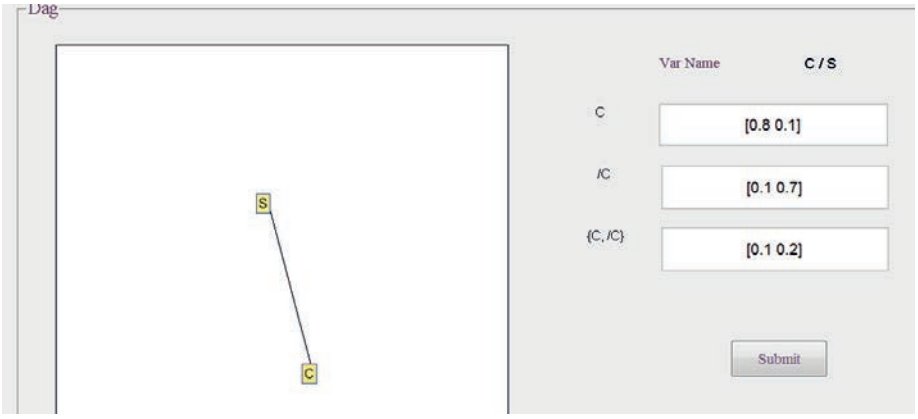


Figure 6.5: Conditional local distribution $m^C(.|s_i)$

6.2.2 Intervening

In order to model causal reasoning and compute the effect of interventions, we have implemented the graph augmentation and graph mutilation methods.

Graph mutilation

For the graph mutilation method, the intervention completely controls the state of the target variable. Accordingly, we have to cut off the links relating the variable concerned by the intervention to its initial causes. Besides, the intervention should not change the beliefs on the direct causes (parents). For that we have to:

- choose the node concerned by the intervention (here the node C),
- delete the edges relating this node to its initial causes,

- change the local mass distribution of the target variable (a certain *bba* focused on the target value here C).

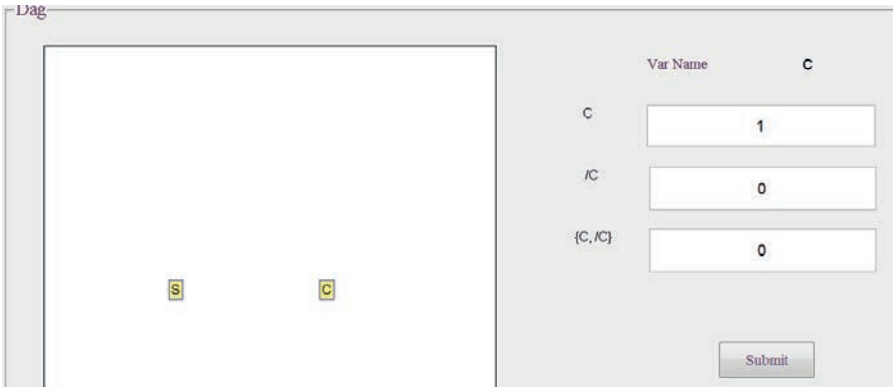


Figure 6.6: Causal belief graph mutilation

Unlike probability and possibility theories, the global joint distribution of networks formalized under the belief function theory are defined on the powerset of the cartesian product of the variables in the network. Note that the computations are based on the uniform ballooning and uniform vacuous extensions that we have defined in Chapter 3.

The global joint distribution of the mutilated graph is presented in Figure 6.7

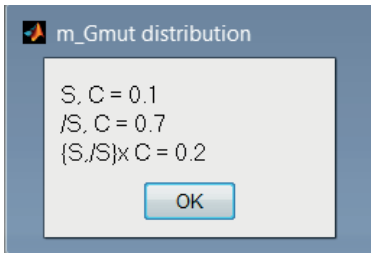


Figure 6.7: $m_{\mathcal{G}_{mut}}$ distribution

Graph augmentation

In this approach, we add a fictive parent node called DO to the variable on which an intervention is performed. The added node is considered as an extra node in the system. This approach allows the computation of the simultaneous effect of observations and interventions.

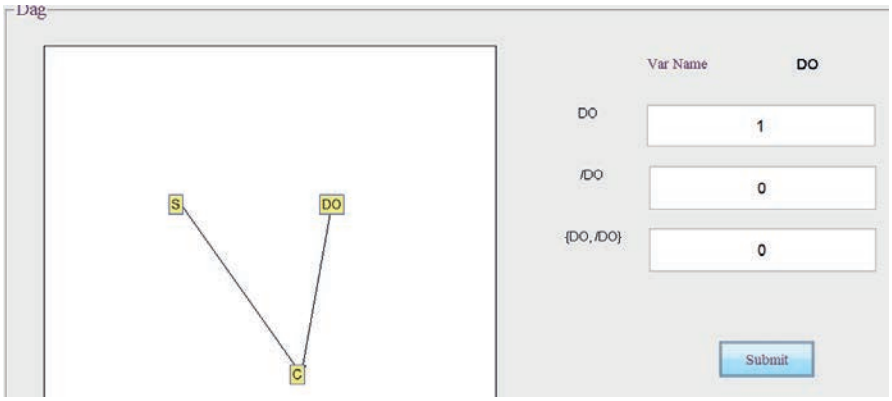


Figure 6.8: Causal belief graph augmentation

The global joint distribution of the augmented graph upon intervening on node C by forcing it to take the value C is given in Figure 6.9.

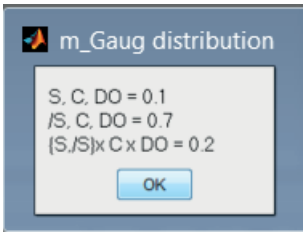


Figure 6.9: $m_{\mathcal{G}_{aug}}$ distribution

From Figure 6.7 and Figure 6.9, we have shown that these different methods to handle interventions are equivalent and have led to the same global belief joint distribution.

6.3 Ascribing causality

6.3.1 Experimental context

With the belief causality ascription model presented in Chapter 5, we propose a method to ascribe causality from:

1. the background knowledge of the agent formalized under the belief function framework;
2. a sequence of reported events (representing potential causes);
3. an abnormal event (the effect) for which we are looking potential causes. Note that causality ascription does not aim to find the actual causes of an event.

Remember that the belief function theory is adequate to handle imperfect causal knowledge and allows to flexibly allocate beliefs to subsets. Besides, the belief causality ascription model allows to handle n-ary variables. Moreover, our method is useful to reason given observations (data directly collected by seeing) and interventions (external manipulations).

For simulation, we have used three scenarios inspired from existing ones (Dickerson & Dickerson, 2000; Halpern & Pearl, 2005; Dubois et al., 2009) but we have modified them to highlight the representation power of our belief model.

- Scenario 1 will present the case where a potential cause is a disjunction of hypotheses;
- Scenario 2 will consider the case of an effect that is a subset of the cartesian product of two variables;
- Scenario 3 will illustrate the usefulness of our model in real world applications through an example in computer security area.

In each scenario, the background knowledge of the agent will be formalized with a causal belief network. With this representation, an agent can simply assess the plausibility that an event occurs in a given context. Variables in the network are either directly observed or manipulated. In each scenario we are trying to answer the question is the reported event A a cause of an abnormal event B?

6.3.2 Creating *bba*s

The belief causality ascription model is built from the agent's background knowledge that is represented with a belief network with conditional beliefs that we have implemented in Section 2. In this network, a priori distributions are uncertain. Uncertainty is represented by *bba*s defined on the set of each variable's instances in the context of its parents. Since the scope of this thesis is modeling and not propagation, we are not able to make inference and compute a posteriori distributions.

Nevertheless, to ascribe the causes of an abnormal event, an agent should have the a posteriori conditional plausibility distribution of the occurrence of the effect in the context of an observed event belonging to the sequence of reported facts. Since we do not have tools to infer such distributions, in the following, we will assume that propagation was made and consequently we have the conditional distribution of the abnormal event given each potential cause. This conditional *bba* is created artificially. So, the question is how will we construct these *bba*s?

Although the effect is an abnormal event, it has taken place after a set of reported events. Accordingly, its occurrence should be plausible. A non-zero plausibility came either from the fact that the effect is a focal element, i.e. has a non-zero *bbm* or one or more subsets that do not contradict with the effect are focal elements, i.e., having a non-zero *bbm*.

Example 6.1. *Let us consider the bbas and their corresponding plausibility distribution defined on* $\Theta = \{\theta_1, \theta_2, \theta_3\}$ *presented in Table 6.1. Assume that the abnormal event is* $\{\theta_1\}$.

Table 6.1: Example of *bba*s with their corresponding *pl*

	m_1	pl_1	m_2	pl_2	m_3	pl_3
$\{\theta_1\}$	0	0	0.7	0.8	0	0.7
$\{\theta_2\}$	0.2	0.3	0.1	0.3	0.2	0.9
$\{\theta_1, \theta_2\}$	0	0.2	0.1	1	0.4	0.9
$\{\theta_3\}$	0.7	0.8	0	0.1	0.1	0.4
$\{\theta_1, \theta_3\}$	0	0.7	0	0.9	0	0.7
$\{\theta_2, \theta_3\}$	0.1	1	0.1	0.3	0	1
$\{\Theta\}$	0	1	0	1	0.3	1

- *In the first bba, m_1, the bbm allocated to $\{\theta_1\}$ and to all its superset is equal to zero. We notice that in this case the occurrence of $\{\theta_1\}$ is not plausible.*
- *In the second bba, m_2, $\{\theta_1\}$ is a focal element. Hence it is plausible.*
- *In the third bba, m_3, $\{\theta_1\}$ is not a focal element. However, some of its supersets namely $\{\theta_1, \theta_2\}$ and Θ are focal elements. In this case, the occurrence of $\{\theta_1\}$ is plausible.*

Consequently, the *bbm* allocated to the abnormal effect or to a subset not contradicting the effect in context of the observed event should be equal to x where x takes value in]0,1]. The value x is generated randomly. In the next subsection, we will explain in details how to generate *bba*s according to different experimental strategies.

6.3.3 Experimental strategy

To check the feasibility of our belief causality ascription model, we will perform several tests and investigate if an observed event from the sequence or reported facts is ascribed as a cause of the effect. Each presented scenario will be used to highlight the advantages of our model. For each scenario, we have made the following tests and finally decide if reported events are ascribed as causes.

Test 1: what if the conditional *bbm* assigned to the effect is non-zero?

In this case, we have generated a conditional *bba* that will be transformed to a conditional plausibility distribution with the Möbius transformation. We make 90 simulations, where each time the *bba* is defined as follows:

1. the *bbm* of the abnormal event is x, $x \in]0, 1]$. We consider several degrees of uncertainty such that:
   - the first 30 simulations: $0 < x \leq 0.25$
   - the second 30 simulations: $0.25 < x \leq 0.5$
   - the third 30 simulations: $0.5 < x \leq 1$
2. *bbm*s of the other subsets are randomly assessed such that the overall sum assigned to these subsets is equal to $(1 - x)$.

For each simulation, we will investigate if the observed event is ascribed as a very strong, strong or weak cause of the abnormal event or not.

Test 2: what if conditional *bbm*s assigned to subsets not contradicting the effect are non-zero?

In this test, we assume that the occurrence of the consequence in context of an observed event is not a focal element. However, one or more subsets that do not contradict with the effect are focal elements, i.e., having a non-zero *bbm*.

We make 90 simulations, where each time the conditional *bba* concerning the abnormal event is defined as follows:

1. the *bbm* of the abnormal event is 0;
2. the *bbm* of a subset not contradicting the abnormal event (generated randomly) is x, $x \in]0, 1]$. We consider several degrees of uncertainty such that:
   - the first 30 simulations: $0 < x \leq 0.25$
   - the second 30 simulations: $0.25 < x \leq 0.5$
   - the third 30 simulations: $0.5 < x \leq 1$
3. *bbm*s of the other subsets are randomly assessed such that the overall sum assigned to these subsets is equal to $(1 - x)$.

After computing the corresponding plausibility of these *bba*s, we apply our belief model to identify whether the observed event is ascribed as a very strong, strong or weak cause of the abnormal event or not.

Test 3: what if *bba*s of all simulations were aggregated?

After running the 90 simulations of test 1 and test 2, an agent can decide about the strength of the cause. For that, the most common way is to compute the percentage of different strengths and choose the highest one.

On the other hand, one can consider that having different distributions is somewhat similar to having several experts such that each expert has provided his own beliefs. The belief function theory offers a tool to aggregate beliefs, namely Dempster's rule of combination.

After aggregating the different distributions, the agent will have to ascribe causality at a final stage. In fact, it is only on the combined distribution that the agent will ascribe causality. Hence through this test, we can compare the results obtained from computing the different rate of each potential strength with those found after merging beliefs using Dempster's rule of combination.

6.3.4 Scenario 1

Let us consider the following sequence of observations: "Anne was asleep under an apple tree. Many children were playing beside her. A child fell on his back. Many apples were ripe. An apple fell on her. She woke up".

Let us suppose that the background knowledge of an agent is given by means of a belief network. As mentioned in Chapter 3, conditional distributions can be defined given any number of parent nodes. As depicted in Figure 6.10, variables in the network are:

- W (for waking up)
  Θ_W={w_1: waking up, w_2: wake up and go back to sleep, w_3: asleep};
- S (for sleeping under an apple tree)
  Θ_S={s_1: yes, s_2: no});
- R (for an apple being ripe)
  Θ_R={r_1: yes, r_2: no};
- F (for an apple falling)
  Θ_F={f_1: yes, f_2 :no};
- C (for children making noise when playing)
  Θ_C={c_1: lot, c_2: little, c_3: no}.
- CF (for a child fell on his back)
  Θ_{CF}={cf_1: yes, cf_1: no}.

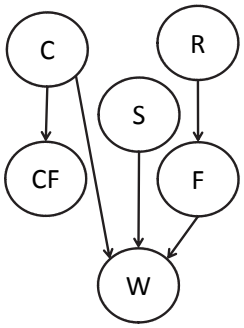


Figure 6.10: Network of scenario 1

Abnormal event and potential causes of scenario 1

In this scenario, waking up (w_1) is an abnormal situation (i.e., a very strongly rejected event) in context of sleeping under an apple tree. We want to ascribe the causes that make Anne wakes up. Potential causes are:

- ca_1: an apple being ripe and falls on her, i.e., $\{(r_1, f_1)\}$;
- ca_2: children making noise when playing, i.e., $\{c_1, c_2\}$;
- ca_3: a child falls on his back, i.e., $\{cf_1\}$.

Notice that the first potential cause ca_1 is an event composed of two variables: being ripe R and fall F.
The second potential cause ca_2 is a case specific to the belief function theory. In fact, the evidence is reported on a disjunction of hypotheses. Here it concerns the subset children making noise $\{c_1, c_2\}$ of the variable C.

Test 1: conditional *bbm* assigned to the effect is non-zero

In this test, we want to know if the fact that the apple is ripe and falls on Anne (ca_1) is the cause of her awakening (w_1) in context of sleeping under an apple tree (s_1). For that we have assigned a conditional *bbm* equal to x to w_1. The value of x has been varied according to three levels as explained in the experimental strategy subsection, namely under 0.25, between 0.25 and 0.5 and greater than 0.5. Figure 6.11 depicts the results obtained after making the 90 simulations.

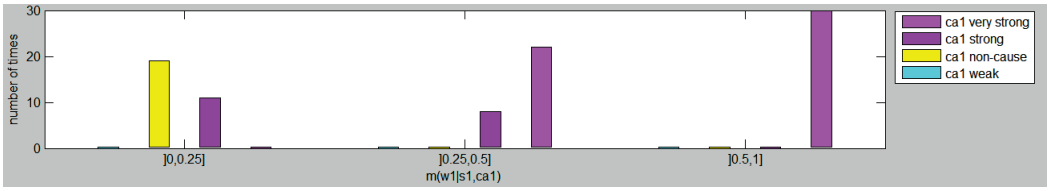


Figure 6.11: Scenario 1, test 1: Is ca_1 a cause of w_1 ?

Table 6.2 summarizes results presented in Figure 6.11. Actually, ca_1 is more often ascribed as a very strong cause.

Table 6.2: Scenario 1, test 1: state of ca_1 per level of uncertainty

$m(w_1\|s_1, ca_1)$	state of the ca_1	number of times	%
]0,0.25]	very strong	0	0
	strong	11	37
	weak	0	0
	non-cause	19	63
]0.25,0.5]	very strong	22	73
	strong	8	27
	weak	0	0
	non-cause	0	0
]0.5,1]	very strong	30	100
	strong	0	0
	weak	0	0
	non-cause	0	0

Generally as shown in Table 6.3, if Anne wakes up then we are in a good position to say that the fact that the apple was ripe and fell on her is a very strong cause of her awakening. In fact, in 58% of the cases ca_1 is ascribed as a very strong cause of w_1.

Table 6.3: Scenario 1, test 1: state of ca_1

	very strong	strong	weak	non-cause
number of times	52	19	0	19
%	58	21	0	21

From Table 6.2 and Table 6.3, we notice that the highest the confidence in the occurrence of the consequence namely w_1 in context of ca_1 is, the strongest ca_1 will be ascribed as a cause of w_1. In particular, we find that when $m(w_1|s_1, ca_1)$ is:

- greater than 0.5 then in 100% of the cases ca_1 is ascribed as a very strong cause of w_1 in context of sleeping under the apple tree.

- between 0.25 and 0.5 then ca_1 is either ascribed as a very strong cause (73% of the cases) or a strong cause (27% of the cases).

- less than 0.25 then ca_1 is either ascribed as a strong cause (37% of the cases) or as a non cause of w_1 (63% of the cases).

Note that weak causes have not been identified here. Indeed, a cause will be identified as a weak cause in a less frequent cases, i.e., if the plausibility of $pl(w_1|s_1, ca_1)$ is either equal to $pl(w_2|s_1, ca_1)$ or $pl(w_3|s_1, ca_1)$.

Test 2: conditional *bbm*s assigned to subsets not contradicting the effect are non-zero

A question that we thought important to investigate under the belief function theory is: "After the propagation, does a *bbm* of 0 assigned to the singleton w_1 in the context of ca_1 suggests that ca_1 is not a cause of w_1"?

Actually unlike probability theory, the plausibility of w_1 in a given context is not only computed from the *bbm* of w_1 in that context. In fact, all *bbm*s assigned to subsets that do not contradict w_1 will be used to compute the plausibility of w_1. In this test, we consider subsets that do not contradict with w_1 (i.e., $\{w_1, w_2\}$, $\{w_1, w_3\}$ and $\{w_1, w_2, w_3\}$) as focal elements. As explaining in the following, we can consider two different situations.

Case of total ignorance: focal elements are $\{w_1, w_2, w_3\}$ It is the case when a *bbm* of 1 is assigned to Θ_W in context of ca_1 (i.e., a vacuous *bba*). In that case, the plausibility of all events is equal to one. Thus, according to our model ca_1 is not ascribed as a cause of w_1.

Case of partial ignorance
This case includes two situations as follows:

1. focal elements are all the supersets of w_1: in this situation, the plausibility of w_1 given ca_1 will be always equal to one. It is also the case of its complement $\{w_2, w_3\}$ given ca_1. Consequently, we are in a situation of ignorance and cannot say that ca_1 is either a cause or not of w_1.

2. at least one focal element is a superset of w_1: it is the general case described in the experimental strategy subsection. Note that the superset of w_1 denoted by w should be different from the frame of discernment. the conditional *bbm* of w in context of ca_1 is equal to x; $x \in]0, 1]$. The value of x has been varied according to three levels, namely under 0.25, between 0.25 and 0.5 and greater than 0.5.

   Figure 6.12 shows the results obtained after making 90 simulations (30 for each level). Table 6.4 and Table 6.5 summarize the results of Figure 6.12.

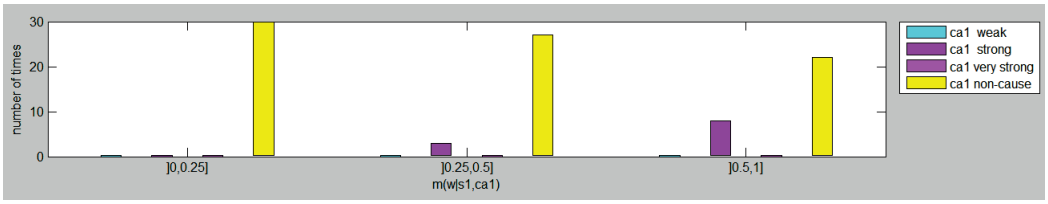


Figure 6.12: Scenario 1, test 2: ascription of ca_1

Table 6.4: Scenario 1, test 2: state of ca_1 per level of uncertainty

$m(w\|s_1, ca_1)$	state of the ca_1	number of times	%
]0,0.25]	very strong	0	0
	strong	0	0
	weak	0	0
	non-cause	30	100
]0.25,0.5]	very strong	0	0
	strong	3	10
	weak	0	0
	non-cause	27	90
]0.5,1]	very strong	0	0
	strong	8	17
	weak	0	0
	non-cause	22	73

From these tables, we can conclude that if Anne wakes up and then perhaps goes back to sleep then we cannot say that the apple was ripe and fell on her is a cause of her awakening. In fact, in 88% of the cases ca_1 is ascribed as a non-cause of w_1.

Table 6.5: Scenario 1, test 2: state of ca_1

	very strong	strong	weak	non-cause
number of times	0	11	0	79
%	0	12	0	88

From Table 6.4 and Table 6.5, we can conclude that:

- if w which is a superset of w_1 in context of ca_1 is a focal element, then it does not mean that ca_1 will be ascribed as a cause of w_1. In fact as shown in Table 6.4, in most cases (88% of the cases) whatever the level of uncertainty, ca_1 is not ascribed as a cause of w_1.
- the percentage where ca_1 is not ascribed as a cause of w_1 depends on the confidence in w in context of ca_1. Thus, it decreases from 100% when it is less than 0.25, to 90% when it is between 0.5 and 0.25 and to 73% if it is greater to 0.5.
- if ca_1 is ascribed as a cause of w_1, it is only a strong cause (12% of the cases). The more confident we are in w in context ca_1, the highest will be the plausibility

that ca_1 is a cause of w_1. In fact, when this *bbm* is between 0.5 and 0.25 then in 10% of the cases ca_1 is ascribed a strong cause of w_1. This rate increases to 17% if $m(w|s_1, ca_1)$ is greater than 0.5.

Test 3: aggregate *bba*s of all simulations

As explained in the experimentation strategy subsection, from the different simulations for a given test an agent can decide about the strength of the cause. The most common way is to compare between the different resulting frequencies.

Having different distributions can be seen as having several experts that have expressed their beliefs. Dempster's rule of combination is therefore used to merge these beliefs into a single distribution. It is only at this final stage that the agent ascribes causality on the aggregated distribution. Accordingly, we will compare the results obtained from computing and comparing the different percentages of each strength of a given cause with those found after merging beliefs using Dempster's rule of combination.

Decision about ca_1 after test 1 In test 1 as mentioned in Table 6.2, we have found that when $m(w_1|s_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is ascribed as a very strong cause.
-]0.5,1], then ca_1 is ascribed as a very strong cause.

From Table 6.3, we can conclude that in general ca_1 is ascribed as a very strong cause of w_1.

By applying Dempster's rule of combination, we have obtained a single distribution. It is on this distribution that we have compared plausibilities as defined in our causality ascription model. We have found that when $m(w_1|s_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is ascribed as a very strong cause.
-]0.5,1], then ca_1 is ascribed as a very strong cause.

By combining all beliefs from the 90 simulations, we have also found that ca_1 is a very strong of w_1.

Decision about ca_1 after test 2 In test 2 as mentioned in Table 6.4, we have found that when $m(w|s_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is a non-cause.
-]0.5,1], then ca_1 is a non-cause.

From Table 6.5, we can conclude that in general ca_1 is a non-cause of w_1.

By applying Dempster's rule of combination, we have found that when $m(w|s_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is ascribed as a non-cause cause.
-]0.5,1], then ca_1 is ascribed as a non-cause.

By combining all beliefs from the 90 simulations we have also found that ca_1 is a non-cause of w_1.

Consequently, both alternatives to decide about the state of the observed event, i.e., if it is not a cause or its strength, have led to the same conclusions.

Discriminate between potential causes

Previously, we have seen how to ascribe one cause. However, in real world we are usually faced to a long sequence of reported events.

In this scenario, Anne woke up. Before that, we have observed three events ca_1: an apple was ripe and fell on her, ca_2: children made noise, ca_3: a child fell on his back.
We want to know if ca_1, ca_2 or ca_3 are potential causes of her awakening. Since we are not looking for the actual cause, we have each time answered the question: is ca_i a cause of w_1? Accordingly, on the basis of the background knowledge we will first compute the plausibility that the consequence namely w_1 occurs in the context of ca_1 as done in subsection 6.3.4. We repeat the same tests in context of ca_2 and ca_3. Remember that ca_2 is a disjunction of hypotheses and this case can be naturally handled under the belief function theory.

Test 1: Under the same experimental condition, reported results about the number of times where ca_1, ca_2 or ca_3 are ascribed as causes of w_1 in context s_1 are shown in Figure 6.13.

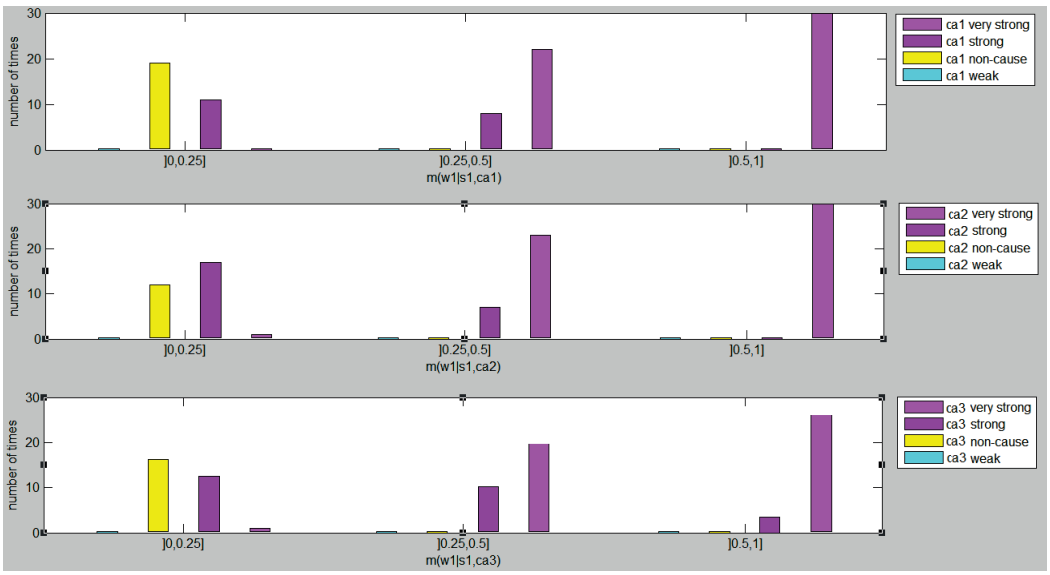


Figure 6.13: Scenario 1, test 1: ascribing causes of w_1

We have already investigated the case where ca_1 is a potential cause of w_1. Thus, we will deal now with ca_2 and ca_3. Table 6.6 represented the results found from the different simulations where $m(w_1|s_1, ca_i)$ is a non-zero *bbm*.

Table 6.6: Scenario 1, test 1: state of ca_i per level of uncertainty

		ca_2		ca_3	
$m(\{w_1\}\|s_1, ca_i)$	state of the ca_1	number	%	number	%
]0,0.25]	very strong	1	3	1	3
	strong	17	57	13	43
	weak	0	0	0	0
	non-cause	12	40	16	54
]0.25,0.5]	very strong	23	77	20	67
	strong	7	23	10	34
	weak	0	0	0	0
	non-cause	0	0	0	0
]0.5,1]	very strong	30	100	25	84
	strong	0	0	5	16
	weak	0	0	0	0
	non-cause	0	0	0	0

Generally as shown in Table 6.7, if Anne wakes up and then we are in a good position to say that children playing and making noise beside her is a cause of her awakening. In fact, in 60% of the cases ca_2 is ascribed as a very strong cause of w_1 and in 27% of the cases it is identified as strong cause. The fact that a child fell on his back is also a potential cause of her awakening. However, it is seen as a weaker cause than ca_2. In fact, in 51% of the cases it is seen as a very strong cause and in 31% of the cases it is ascribed as a strong cause.

Table 6.7: Scenario 1, test 1: state of ca_2 and ca_3

	ca_i	very strong	strong	weak	non-cause
number of times	ca_2	54	24	0	12
	ca_3	46	28	0	16
%	ca_2	60	27	0	13
	ca_3	51	31	0	18

Results obtained from Table 6.6 and Table 6.7 concerning ca_2 et ca_3 confirm those found when dealing with ca_1. In fact, if the confidence in the occurrence of the consequence namely w_1 in context of ca_2 increases, ca_2 will be ascribed as a stronger cause of w_1. Indeed, if the *bbm* allocated to w_1 in context of the potential cause is:

- greater than 0.5 then in most cases (100% of the cases for ca_2 and 84% of the cases for ca_3), the observed event is ascribed as a very strong cause of w_1 in context of sleeping under the apple tree.

- between 0.25 and 0.5 then ca_i is either ascribed as a very strong cause (77% of the cases for ca_2 and 67% of the cases for ca_3) or a strong cause (23% of the cases for ca_2 and 34% of the cases for ca_3).

- less than 0.25 then ca_i is generally ascribed as either as a strong cause (57% of the cases for ca_2 and 43% of the cases for ca_3) or as a non cause of w_1 (40% of the cases for ca_2 and 54% of the cases for ca_3).

Test2: We work under the same experimental condition, i.e., w_1 is not a focal element. Reported results about the number of times where ca_1, ca_2 or ca_3 are ascribed as causes of w_1 in context s_1 are shown in Figure 6.14.

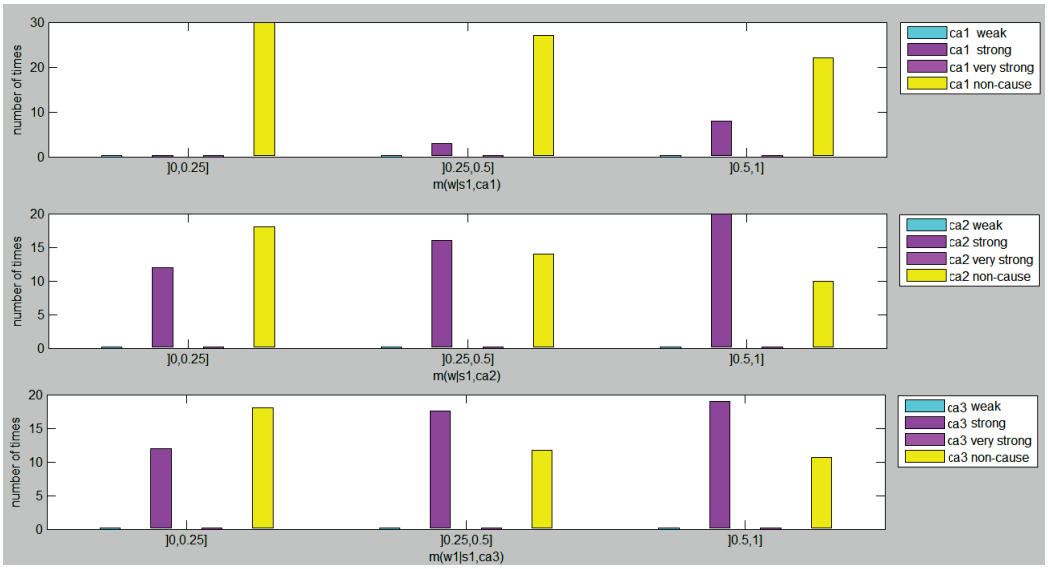


Figure 6.14: Scenario 1, test 2: ascription of ca_1, ca_2 and ca_3

Based on Figure 6.14, Table 6.8 represents the results found from the different simulations where $m(w|s_1, ca_i) > 0$ such that w is a superset of w_1.

Table 6.8: Scenario 1, test 2: state of ca_i per level of uncertainty

		ca_2		ca_3	
]0,0.25]	very strong	0	0	0	0
	strong	12	40	12	40
	weak	0	0	0	0
	non-cause	18	60	18	60
]0.25,0.5]	very strong	0	0	0	0
	strong	16	53	17	57
	weak	0	0	0	0
	non-cause	14	47	13	43
]0.5,1]	very strong	0	0	0	0
	strong	20	67	19	63
	weak	0	0	0	0
	non-cause	10	33	11	37

As shown in Table 6.9, if Anne wakes up and then perhaps goes back to sleep then the fact that children make noise beside her is either considered as a strong cause or a non-cause of her awakening. It is the same case when observing a child falling on his back.

Table 6.9: Scenario 1, test 2: state of ca_2 and ca_3

		very strong	strong	weak	non-cause
number of times	ca_2	0	48	0	42
	ca_3	0	48	0	42
%	ca_2	0	53	0	47
	ca_3	0	53	0	47

From Table 6.8 and Table 6.9, we can conclude that:

- the percentage where ca_1 is not ascribed as a cause of w_1 depends on the confidence in w in context of ca_i. Thus, it decreases from 60% for ca_2 and ca_3 when it is less than 0.25, to 47% for ca_2 and to 43% for ca_3 when it is between 0.5 and 0.25 and to 33% for ca_2 and to 37% for ca_3 if it is greater to 0.5.

- if ca_i is ascribed as a cause of w_1, it is only a strong cause (53% of the cases). The more confident we are in w in context ca_i, the highest will be the plausibility that ca_i is a cause of w_1. In fact, when this *bbm* is less than 0.25 then in 40% of the cases ca_2 is ascribed as a strong cause of w_1 and in 40% of the cases ca_3 is ascribed as a strong cause of w_1. This rate increases to 53% if $m(w|s_1, ca_2)$ is between 0.5 and 0.25 and to 44% if $m(w|s_1, ca_3)$ is between 0.5 and 0.25. It becomes respectively 67% and 63% if the *bbm* is greater than 0.5.

Test 3:

- Decision about ca_2 after test 1:
  In Test 1, we have found that when $m(w_1|s_1, ca_2)$ is in:
  -]0,0.25], then ca_2 is ascribed as a strong cause.
  -]0.25,0.5], then ca_2 is ascribed as a very strong cause.
  -]0.5,1], then ca_2 is ascribed as a very strong cause.
  From Table 6.7, we can conclude that in general ca_2 is ascribed as a very strong cause of w_1.

  By applying Dempster's rule of combination, we have found that when $m(w_1|s_1, ca_2)$ is in:
  -]0,0.25], then ca_2 is ascribed as a strong cause.
  -]0.25,0.5], then ca_2 is ascribed as a very strong cause.
  -]0.5,1], then ca_2 is ascribed as a very strong cause.
  By combining all beliefs from the 90 simulations, we have also found that ca_2 is a very strong of w_1.

- Decision about ca_2 after test 2:
  In Test 2, we have found that when $m(w|s_1, ca_2)$ is in:
  -]0,0.25], then ca_2 is a non-cause.
  -]0.25,0.5], then ca_2 is ascribed as a strong cause.
  -]0.5,1], then ca_2 is ascribed as a strong cause.
  From Table 6.9, we can conclude that in general ca_2 is a strong cause of w_1.

By applying Dempster's rule of combination, we have found that when $m(w|s_1, ca_2)$ is in:
-]0,0.25], then ca_2 is a non-cause.
-]0.25,0.5], then ca_2 is ascribed as a strong cause.
-]0.5,1], then ca_2 is ascribed as a strong cause.

By combining all beliefs from the 90 simulations we have also found that ca_2 is a strong cause of w_1.

The two alternatives used to decide about the state of the observed event ca_2 have led to the same conclusions.

- Decision about ca_3 after test 1:
  In Test 1, we have found that when $m(w_1|s_1, ca_3)$ is in:
  -]0,0.25], then ca_3 is ascribed as a non-cause cause.
  -]0.25,0.5], then ca_3 is ascribed as a very strong cause.
  -]0.5,1], then ca_3 is ascribed as a very strong cause.

  From Table 6.7, we can conclude that in general ca_3 is ascribed as a very strong cause of w_1.

  By applying Dempster's rule of combination, we have found that when $m(w_1|s_1, ca_3)$ is in:
  -]0,0.25], then ca_3 is a non-cause cause.
  -]0.25,0.5], then ca_3 is ascribed as a very strong cause.
  -]0.5,1], then ca_3 is ascribed as a very strong cause.

  By combining all beliefs from the 90 simulations, we have also found that ca_3 is a very strong cause of w_1.

- Decision about ca_3 after test 2:
  In Test 2, we have found that when $m(w|s_1, ca_3)$ is in:
  -]0,0.25], then ca_3 is a non-cause.
  -]0.25,0.5], then ca_3 is ascribed as a strong cause.
  -]0.5,1], then ca_3 is ascribed as a strong cause.

  From Table 6.9, we can conclude that in general ca_3 is a strong cause of w_1.

  By applying Dempster's rule of combination, we have found that when $m(w|s_1, ca_3)$ is in:
  -]0,0.25], then ca_3 is a non-cause.
  -]0.25,0.5], then ca_3 is ascribed as a strong cause.
  -]0.5,1], then ca_3 is ascribed as a strong cause.

  By combining all beliefs from the 90 simulations we have also found that ca_3 is a strong cause of w_1.

Here also the two alternatives used to decide about the state of the observed event ca_3 have led to the same conclusions.

In this scenario, we have pointed out that our model can handle n-ary variables. Besides, we have shown that causes can be a disjunction of hypotheses. In fact, the observed event ca_2 corresponding to the event children making noise is represented with the subset $\{c_1, c_2\}$ reflecting the hesitation between saying that children have made a lot or little noise.

6.3.5 Scenario 2

Let us consider the following sequence of observations: "We were at a forest, the wood was dry. There was a lightening. An arsonist was with us playing with a match. A forest fire and a big house fire started".

A belief network represent the background knowledge of an agent. Figure 6.15 shows the relation between variables in the network.

- H (for house fire)
  Θ_H={h_1: big, h_2: medium, h_3: small, h_4: no};
- F (for forest fire)
  Θ_F={f_1: yes, f_2: no};
- L (for lightening)
  Θ_L={l_1: yes, l_2: no};
- D (for wood dry)
  Θ_D={d_1: yes, d_2: no};
- M (for light a match)
  Θ_M={m_1: yes, m_2: no}.

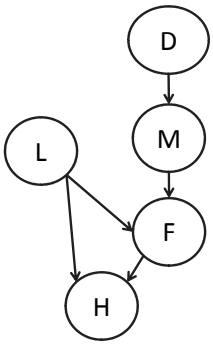


Figure 6.15: Network of scenario 2

Abnormal event and potential causes of scenario 2

A forest fire and a big house fire is an abnormal event denoted by e, in context of wood dry. With our belief model, we can consider that the consequence is the abnormal event $e \subseteq \Theta_H \times \Theta_F$, such that e is composed of a conjunction of big house fire ($\{h_1\}$) and forest fire (which is a disjunction of all kinds of forest fires $\{f_1, f_2\}$). Hence, $e = \{f_1, f_2\} \times \{h_1\}$.

In context of wood dry, we want to ascribe causes of e such that potential causes are:

- ca_1: a lightening, i.e., $\{l_1\}$;
- ca_2: an arsonist lit a match, i.e., $\{m_1\}$.

Test 1: conditional *bbm* assigned to the effect is non-zero

In this test, we want to know if either the lightening (ca_1) or matching a lit (ca_2) is the cause of the forest fire and the big house fire (e). For that we have assigned a conditional *bbm* equal to x to e. The value of x has been varied according to three levels, namely under 0.25, between 0.25 and 0.5 and greater than 0.5.
Figure 6.16 presents the results of causality ascription of ca_1 and ca_2, after making 90 simulations (30 per level). Table 6.10 and Table 6.11 summarize these different results.

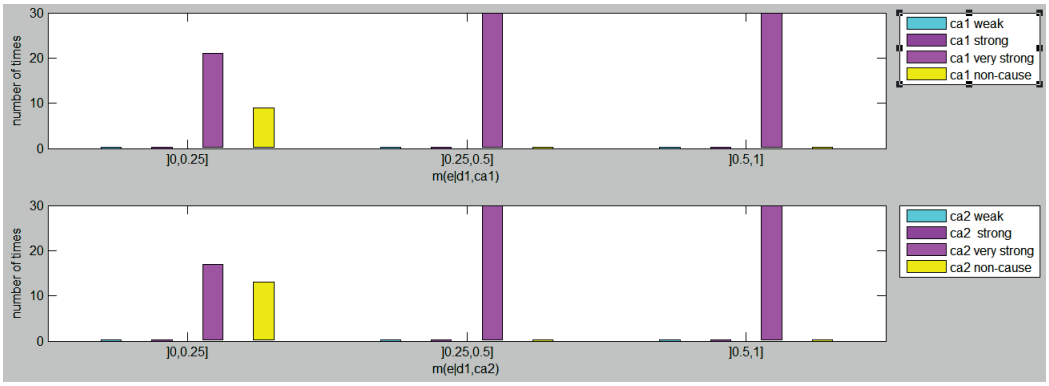


Figure 6.16: Scenario 2, test 1: ascribing causes of e

Table 6.10: Scenario 2, test 1: state of ca_i per level of uncertainty

		ca_1		ca_2	
$m(e\|d_1, ca_i)$	state of the ca_i	number	%	number	%
]0,0.25]	very strong	21	70	17	57
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	9	30	13	43
]0.25,0.5]	very strong	30	100	30	100
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	0	0	0	0
]0.5,1]	very strong	30	100	30	100
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	0	0	0	0

Table 6.11: Scenario 2, test 1: state of ca_1 and ca_2

	ca_i	very strong	strong	weak	non-cause
number of times	ca_1	81	0	0	9
	ca_2	77	0	0	13
%	ca_1	90	0	0	10
	ca_2	86	0	0	14

From these tables, we notice that if the *bbm* allocated the event e in context of the potential cause is:

- greater than 0.25 then in 100% of the cases ca_1 and ca_2 are ascribed as a very strong cause of e.
- less than 0.25 then the observed event is ascribed in most cases as a very strong cause (70% of the cases for ca_1 and 57% for ca_2).

Test 2: conditional *bbm*s assigned to subsets not contradicting the effect are non-zero

As done in scenario 1, we have studied here the case where a *bbm* of zero is assigned to the event e in the context of d_1 and a non-zero *bbm* is assigned to a subset that do not contradict with e. Obviously, as for scenario 1, in the case where a *bbm* of one is assigned to $\Theta_H \times \Theta_F$ or it is shared between only events that do not contradict with e, then we are in a case of ignorance. Consequently, we are not able to say if the observed event is a cause of e.

In the following, we will investigate the case of partial ignorance by considering the case where at least one focal element denoted by e' does not contradict with e. The conditional *bba* defined in context of d_1,ca_1 (resp. d_1, ca_2) is given by:

- the *bbm* assigned to e is 0;
- a random focal element, e', different from the frame of discernment and that do not contradict with e having a *bbm* of x; $x \in]0,1]$. As explained in the experimental strategy subsection, we have considered three levels of uncertainty of x namely x belonging to]0,0.25],]0.25,0.5] and]0.5,1]);
- remaining subsets randomly share a mass of $1-x$.

As shown in Figure 6.17, results of causality ascription depend on the *bbm* of e' in context of d_1.

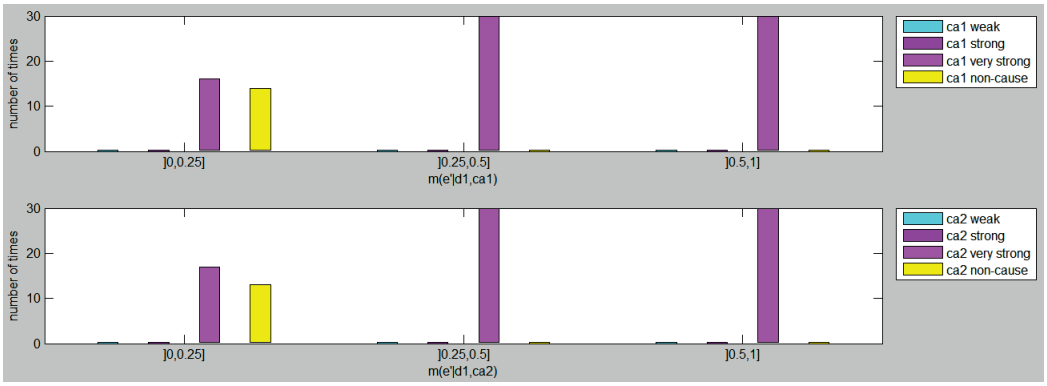


Figure 6.17: Scenario 2, test 2: ascription of ca_1 and ca_2

Table 6.12 summarizes the results of Figure 6.17. Table 6.13 presents the final results about the 90 simulations for each potential cause.

Table 6.12: Scenario 2, test 2: state of ca_i per level of uncertainty

		ca_1		ca_2	
$m(e'\|d_1, ca_i)$	state of the ca_1	number	%	number	%
]0,0.25]	very strong	16	53	17	57
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	14	47	13	43
]0.25,0.5]	very strong	30	100	30	100
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	0	0	0	0
]0.5,1]	very strong	30	100	30	100
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	0	0	0	0

Table 6.13: Scenario 2, test 2: state of ca_1 and ca_2

	ca_i	very strong	strong	weak	non-cause
number of times	ca_1	76	0	0	14
	ca_2	77	0	0	13
%	ca_1	84	0	0	16
	ca_2	86	0	0	14

From these tables, we can conclude that when the *bbm* allocated to e' in context of wood dry and the observed event is:

- greater than 0.5 then in 100% of the cases ca_1 and ca_2 are ascribed as a very strong cause of e.

- less than 0.25 then the observed event is ascribed in most cases as a very strong cause (53% of the cases for ca_1 and 57% for ca_2).

Test 3: aggregate *bba*s of all simulations

Dempster's rule of combination is used to merge the different distributions generated from the 90 simulations. At the end, the agent ascribes causality on the aggregated distribution. In the following, we will compare the results given after ascribing the causes for each simulation with those found after merging beliefs.

Decision about ca_1 after test 1
Let us start with ca_1 when the *bba* is defined according to the requirements of test 1:

As mentioned in Table 6.10, we have found that when $m(e|d_1, ca_1)$ is in:
-]0,0.25], then ca_1 is ascribed as a very strong cause.
-]0.25,0.5], then ca_1 is ascribed as a very strong cause.

-]0.5,1], then ca_1 is ascribed as a a very strong cause.
From Table 6.11, we can conclude that in general ca_1 is ascribed as a very strong cause of e.

By applying Dempster's rule of combination, we have found that when $m(e|d_1, ca_1)$ is in:
-]0,0.25], then ca_1 is ascribed as a very strong cause.
-]0.25,0.5], then ca_1 is ascribed as very strong cause.
-]0.5,1], then ca_1 is ascribed as a very strong cause.
By combining all beliefs from the 90 simulations, we have also found that ca_1 is ascribed as a very strong cause of e.

Decision about ca_1 after test 2
Let us now compare the results about the state of ca_1 when the *bba* is defined according to the requirements of test 2.

As mentioned in Table 6.12, we have found that when $m(e'|d_1, ca_1)$ is in:
-]0,0.25], then ca_1 is ascribed as a very strong cause.
-]0.25,0.5], then ca_1 is ascribed as a very strong cause.
-]0.5,1], then ca_1 is ascribed as a very strong cause.
From Table 6.13, we can conclude that in general ca_1 is ascribed as a very strong cause of e.

By applying Dempster's rule of combination, we have found that when $m(e'|d_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a ascribed as a very strong cause.
-]0.25,0.5], then ca_1 is a ascribed as a very strong cause.
-]0.5,1], then ca_1 is a ascribed as a very strong cause.
By combining all beliefs from the 90 simulations, we have also found that ca_1 is a ascribed as a very strong cause of e.

Decision about ca_2 after test 1

Let us consider the results concerning ca_2 when the *bbm* satisfies the requirement of test 1. As it is presented in Table 6.10 that when $m(e|d_1, ca_2)$ is in:
-]0,0.25], then ca_2 is ascribed as a very strong cause.
-]0.25,0.5], then ca_2 is ascribed as a very strong cause.
-]0.5,1], then ca_2 is ascribed as a very strong cause.
From Table 6.11, we can conclude that in general ca_2 is ascribed as a very strong cause of e.

By applying Dempster's rule of combination, we have found that when $m(e|d_1, ca_2)$ is in:
-]0,0.25], then ca_2 is ascribed as a very strong cause
-]0.25,0.5], then ca_2 is ascribed as a very strong cause.
-]0.5,1], then ca_2 is ascribed as a very strong cause.
By combining all beliefs from the 90 simulations, we have also found that ca_2 is ascribed as a very strong cause of e.

Decision about ca_2 after test 2
Let us now compare the results about the state of ca_2 when the *bba* is defined according to

the requirements of test 2.

As mentioned in Table 6.12, we have found that when $m(e'|d_1, ca_2)$ is in:
-]0,0.25], then ca_2 is ascribed as a very strong cause.
-]0.25,0.5], then ca_2 is ascribed as a very strong cause.
-]0.5,1], then ca_2 is ascribed as a very strong cause.

From Table 6.13, we can conclude that in general ca_2 is ascribed as a very strong cause of e.

By applying Dempster's rule of combination, we have found that when $m(e'|d_1, ca_2)$ is in:
-]0,0.25], then ca_2 is a ascribed as a very strong cause.
-]0.25,0.5], then ca_2 is a ascribed as a very strong cause.
-]0.5,1], then ca_2 is a ascribed as a very strong cause.

By combining all beliefs from the 90 simulations, we have also found that ca_2 is a ascribed as a very strong cause of e.

We notice that the two alternatives used to decide about the state of the observed event either ca_1 or ca_2 have led to the same results.

In this scenario, we have pointed out that our model can deal with consequences that are subsets of the cartesian product of several variables. In fact, the abnormal event e corresponds to the event a forest fire and a big house fire which is represented with $\{f_1, f_2\} \times \{h_1\}$.

6.3.6 Scenario 3

In this scenario, we have illustrated the use of our model on computer security area.

Let us consider the following sequence of observations: "An attacker conducts a stealthy port scan of a network, sending packets to several well-known ports (e.g., ftp, telnet, http) looking for systems that might be running those services. Those services are present on the system. The attacker penetrate the system."

The background knowledge of an agent is depicted in Figure 6.18. Variables are represented with nodes such that:

- Att (for an attack)
  Θ_{Att}=$\{a_1$: DOS, a_2: $U2R$, a_3: no$\}$;

- Ser (for services)
  Θ_{Ser}=$\{ser_1$: yes, ser_2: no$\}$;

- P (for sending hight number of packets)
  Θ_P=$\{p_1$: yes, p_2: no$\}$;

- Sc (for scan)
  Θ_{Sc}=$\{sc_1$: yes, sc_2: no$\}$).

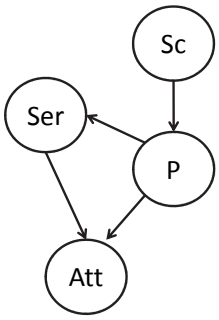


Figure 6.18: Network of scenario 3

Abnormal event and potential causes of scenario 3

A *DOS* attack is an abnormal event in context of running services ser_1. We want to ascribe causes of this attack a_1 such that potential causes are:

- ca_1: packets were sent, i.e., $\{p_1\}$;
- ca_2: a scan was made, i.e., $\{sc_1\}$.

Test 1: conditional *bbm* assigned to the effect is non-zero

In this test, we want to know if either the fact that packets were sent (ca_1) or a scan was made(ca_2) is the cause of a *DOS* attack (a_1). For that we have assigned a conditional *bbm* equal to x to a_1. The value of x has been varied according to three levels, namely under 0.25, between 0.25 and 0.5 and greater than 0.5. Figure 6.19 presents the results of causality ascription for all the potential causes, namely ca_1 and ca_2, after making 90 simulations (30 per level).

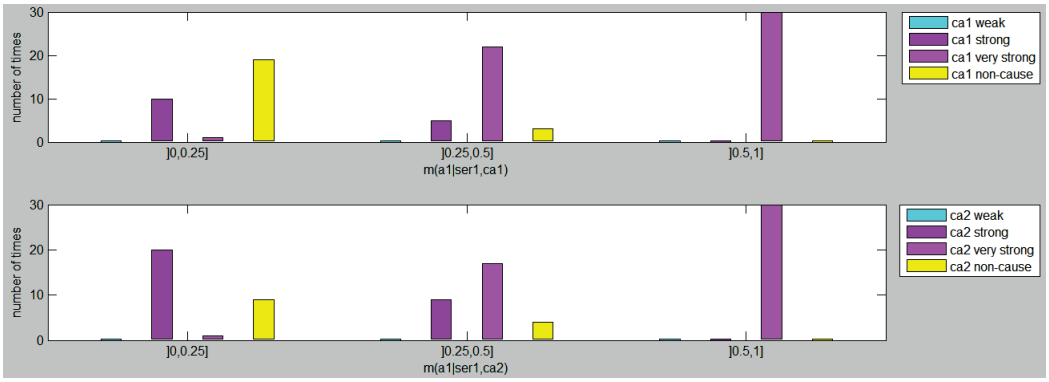


Figure 6.19: Scenario 3, test 1: ascribing causes of a_1

Table 6.14 summarizes the different results presented in Figure 6.19. General results are presented in Table 6.15.

Table 6.14: Scenario 3, test 1: state of ca_i per level of uncertainty

		ca_1		ca_2	
$m(a_1\|ser_1, ca_i)$	state of the ca_1	number	%	number	%
]0,0.25]	very strong	1	3	1	3
	strong	10	33	20	67
	weak	0	0	0	0
	non-cause	19	64	9	30
]0.25,0.5]	very strong	22	73	17	57
	strong	5	17	9	30
	weak	0	0	0	0
	non-cause	3	10	4	13
]0.5,1]	very strong	30	100	30	100
	strong	0	0	0	0
	weak	0	0	0	0
	non-cause	0	0	0	0

Table 6.15: Scenario 3, test 1: state of ca_1 and ca_2

	ca_i	very strong	strong	weak	non-cause
number of times	ca_1	53	15	0	22
	ca_2	48	29	0	13
%	ca_1	59	17	0	24
	ca_2	53	32	0	15

From these tables, we conclude that the highest the confidence in the occurrence of the consequence a_1 is, the strongest the observed event will be ascribed as a cause of a_1. In particular, we find that when it is:

- greater than 0.5 then in 100% of the cases ca_1 and ca_2 are ascribed as a very strong cause of a_1.
- between 0.25 and 0.5 then ca_1 is either ascribed as a very strong cause (73% of the cases for ca_1 and 57% for ca_2) or a strong cause (17% of the cases for ca_1 and 30% for ca_2).
- less than 0.25 then ca_1 is ascribed as a very strong cause is few cases (3% of the cases for ca_1 and ca_2), a strong cause (33% of the cases for ca_1 and 67% for ca_2) or as a non cause of a_1 (64% of the cases for ca_1 and 30% for ca_2).

Test 2: conditional *bbm*s assigned to subsets not contradicting the effect are non-zero

What if we are not certain that an attack took place, i.e., we support $\{a_1, a_2, a_3\}$ or we do not know precisely its type, for instance it can be a DOS or an $U2R$ attack. In this case a *bbm* of zero is assigned to a_i in the context of ser_1 and $\{a_1, a_2\}$ will be a focal element.

As mentioned in the previous scenarios, in the case of a conditional vacuous *bba* (only one focal element Θ_{Att}) then we are in a case of ignorance. Moreover, if all the subsets that do

not contradict a_1 are focal elements then the plausibility of a_1 as well as the plausibility of $\overline{a_1}$ are equal to one. Consequently, in both cases we are not able to say if ca_i is a cause of the *DOS* attack.

In the following, we will investigate the case of partial ignorance where we only know that the attacker succeeds to penetrate the system. Accordingly, the conditional *bba* in context of ser_1, ca_i satisfies these constraints:

- the *bbm* assigned to a_1 is 0;
- the *bbm* of focal element $\{a_1, a_2\}$, reflecting our imprecision about the kind of the attack is x; $x \in]0,1]$. Three levels of uncertainty of x were defined, namely x belonging to]0,0.25],]0.25,0.5] and]0.5,1] are considered;
- remaining elements randomly share a mass of $1 - x$.

For each level of uncertainty, we have generated 30 times our model. Figure 6.20 depicts the different ascribed strength of ca_1 and ca_2 respectively in this case of partial ignorance.

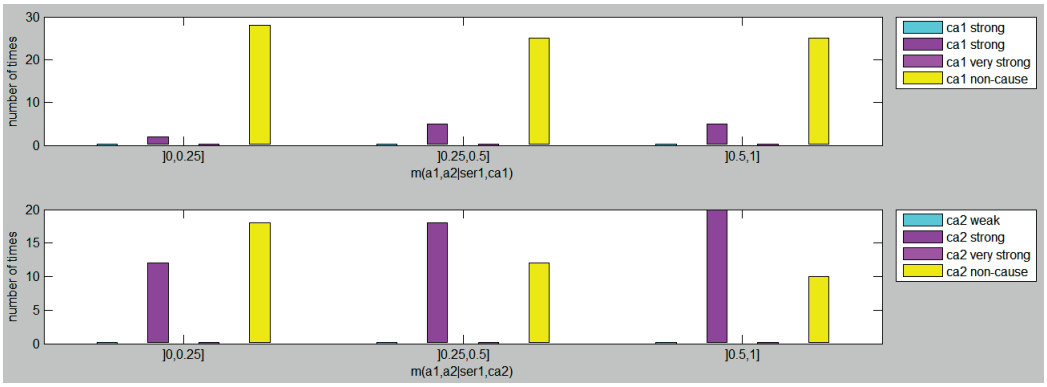


Figure 6.20: Scenario 3, test 2: ascription of ca_1 and ca_2

Table 6.16 summarizes the results of Figure 6.20. Table 6.17 presents the final results about the 90 simulations for ca_1 and ca_2.

Table 6.16: Scenario 3, test 2: state of ca_i per level of uncertainty

		ca_1		ca_2	
$m(\{a_1, a_2\}\|ser_1, ca_i)$	state of the ca_1	number	%	number	%
]0,0.25]	very strong	0	0	0	0
	strong	2	7	12	40
	weak	0	0	0	0
	non-cause	28	93	18	60
]0.25,0.5]	very strong	0	0	0	0
	strong	5	17	18	60
	weak	0	0	0	0
	non-cause	25	83	12	40
]0.5,1]	very strong	0	0	0	0
	strong	5	17	20	67
	weak	0	0	0	0
	non-cause	25	83	10	33

Table 6.17: Scenario 3, test 2: state of ca_1 and ca_2

	ca_i	very strong	strong	weak	no-cause
number of times	ca_1	0	12	0	78
	ca_2	0	50	0	40
%	ca_1	0	13	0	87
	ca_2	0	56	0	44

From these tables, we can conclude that:

- the number of times where ca_1 and ca_2 are identified as non-causes of the *DOS* attack a_1 depends on the confidence in $\{a_1, a_2\}$ in context of ca_1 and on context of ca_2. In fact, for ca_1 it decreases from 93% when it is less than 0.25, to 83% when it is greater than 0.25. We notice the same behavior for ca_2 where it decreases from 60% when it is less than 0.25, to 40% when it is between 0.25 and 0.5 to 33% if it is greater to 0.5.

- if ca_1 or ca_2 is ascribed as a cause of a_1, it is always a strong cause. The most confident we are in $\{a_1, a_2\}$, the highest will be the possibility that the observed event is cause of a_1. In fact,
  - for ca_1, when this *bbm* is less than 0.25 then in 7% of the cases ca_1 is ascribed as a strong cause of a_1. This rate increases to 17% if $m(w|s_1, ca_1)$ is greater than 0.25.
  - for ca_2, when this *bbm* is less than 0.25 then in 40% of the cases ca_2 is ascribed as a strong cause of a_1. This rate increases to 60% if $m(w|s_1, ca_1)$ is between 0.25 and 0.5 and to 67% if it is greater than 0.25

Test 3: aggregate *bba*s of all simulations

As for the two previous scenarios, for each observed event ca_1 and ca_2, we will compare the results of our causality ascription model found after for each simulation with those found after merging beliefs using Dempster's rule of combination.

Decision about ca_1 after test 1
Let us start with ca_1 when the *bba* is defined according to the requirements of test 1:

As mentioned in Table 6.14, we have found that when $m(a_1|ser_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is ascribed as a very strong cause.
-]0.5,1], then ca_1 is ascribed as a a very strong cause.

From Table 6.15, we can conclude that in general ca_1 is ascribed as a very strong cause of a_1.

By applying Dempster's rule of combination, we have found that when $m(a_1|ser_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause
-]0.25,0.5], then ca_1 is ascribed as very strong cause.
-]0.5,1], then ca_1 is ascribed as a very strong cause.

By combining all beliefs from the 90 simulations, we have also found that ca_1 is ascribed as a very strong cause of a_1.

Decision about ca_1 after test 2
Let us now compare the results about the state of ca_1 when the *bba* is defined according to the requirements of test 2.

As mentioned in Table 6.16, we have found that when $m(\{a_1, a_2\}|ser_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is a non-cause.
-]0.5,1], then ca_1 is a non-cause.

From Table 6.17, we can conclude that in general ca_1 is a non-cause of a_1.

By applying Dempster's rule of combination, we have found that when $m(\{a_1, a_2\}|ser_1, ca_1)$ is in:
-]0,0.25], then ca_1 is a non-cause.
-]0.25,0.5], then ca_1 is a non-cause.
-]0.5,1], then ca_1 is a non-cause.

By combining all beliefs from the 90 simulations, we have also found that ca_1 is a non-cause of a_1.

Decision about ca_2 after test 1
Results concerning ca_2 when the*bbm* satisfies the requirement of test 1 are presented in Table 6.14. In fact, when $m(a_1|ser_1, ca_2)$ is in:
-]0,0.25], then ca_2 is ascribed as a strong cause.
-]0.25,0.5], then ca_2 is ascribed as a very strong cause.
-]0.5,1], then ca_2 is ascribed as a very strong cause.

From Table 6.15, we can conclude that in general ca_2 is ascribed as a very strong cause of a_1.

By applying Dempster's rule of combination, we have found that when $m(a_1|ser_1, ca_2)$ is in:
-]0,0.25], then ca_2 is ascribed as strong cause
-]0.25,0.5], then ca_2 is ascribed as very strong cause.

-]0.5,1], then ca_2 is ascribed as a very strong cause.
By combining all beliefs from the 90 simulations, we have also found that ca_2 is ascribed as a very strong cause of a_1.

Decision about ca_2 after test 2
Let us consider the results presented in Table 6.16 about ca_2 when the *bbm* satisfies the requirement of test 2. In fact, when $m(\{a_1, a_2\}|ser_1, ca_2)$ is in:
-]0,0.25], then ca_2 is a non-cause.
-]0.25,0.5], then ca_2 is ascribed as a strong cause.
-]0.5,1], then ca_2 is ascribed as a strong cause.
From Table 6.17, we can conclude that in general ca_2 is ascribed as a strong cause of a_1.

By applying Dempster's rule of combination, we have found that when $m(\{a_1, a_2\}|ser_1, ca_2)$ is in:
-]0,0.25], then ca_2 is a non-cause.
-]0.25,0.5], then ca_2 is a strong cause.
-]0.5,1], then ca_2 is a strong cause.
By combining all beliefs from the 90 simulations, we have also found that ca_2 is a strong cause of a_1.

We notice that the two alternatives used to decide about the state of the observed event either ca_1 or ca_2 have led to the same results.

In this scenario, we have shown the use of our model in context of the computer security area to discriminate between potential causes of an attack. Even though, belief networks have been exploited in many real-world applications (e.g., threat assessment (Benavoli et al., 2009), system analysis (Simon et al., 2008)), it will be difficult to use our approach in some applications requiring in their modeling a high number of nodes. Indeed, the major drawback of the belief function theory remains its high computational complexity since it works on subsets of the frame of discernment.

6.4 Conclusion

In this chapter, we first described the developed causal belief network on which we simulated the effect of an intervention using the belief graph mutilation and augmentation approaches. Indeed, on this network we are able to compute the simultaneous effect of observations and interventions.
Then, we presented several scenarios to show the feasibility and the usefulness of the belief causality ascription model to ascribe the causes of an abnormal situation from a sequence of reported events.

Conclusion

This PhD thesis was devoted to the modeling of causal reasoning from uncertain knowledge. Uncertainty is formalized under the belief function framework.

Causality should be well distinguished from spurious correlation. For that we have focused on the concept of interventions under the belief function framework. Interventions are external manipulations to the system that force a target variable to have a specific value. An action on a cause forcing it to take place will lead to the occurrence of the effect. As for probability and possibility theories, a graphical structure (a causal network) is needed to specify the mechanisms concerned by these external actions.

Besides, a causal network can be seen as an associational network where arcs do not only represent dependence relations. More precisely, the parent set of a given variable is seen as its immediate causes. Accordingly, the equivalence hypothesis of directed associational networks is not valid any more. Thus, a causal network is a proper associational network but the contrary is not always true. This means that its structure is more meaningful and more expressive than the standard network.

For that reasons, as a main contribution we have revised the theoretical foundations of existing belief networks and we have proposed a new associational belief network so-called Belief network with conditional beliefs (BNC). This latter is a directed acyclic graph where conditional *bba*s can be defined in the context of one or more than one parent nodes. To compute the global joint distribution, we have proposed a new belief chain rule based on uniform operations to get rid from conditioning and to extend beliefs to a joint frame.

Another main contribution of this thesis is a graphical model to deal with interventions namely the causal belief network with conditional beliefs ($CBNC$). This network is an appropriate model to represent imperfect causal knowledge in particular it is ideal for ignorance situations. It consists of a BNC with a special interpretation in the sense that edges point from causes toward effects.

We have presented an approach to define belief causation and model the effect of interventions. Therefore, a generalization of the "do" operator under the belief function framework was proposed to compute the effect of these external actions. We have shown that interventions are graphically handled by altering the structure of the causal graph. Mutilation

consists of cutting off the link relating the variable of interest to its parents, whereas the second method, namely augmentation, consists of adding a fictive node DO representing the intervention as a new parent of the manipulated variable. Even if the joint distribution is not defined as for probability distribution, these two different methods remain equivalent under the belief function framework.

We have also investigated the case of non-standard external actions. Indeed, we have pointed out that an intervention can be imperfect or have imperfect consequences. By imperfect interventions, we refer to interventions that may for instance uncertainly occur (i.e., takes place with a degree of belief). An intervention can have imperfect consequences in the sense that it does not succeed to put its target into one specific value. We have shown that these kinds of interventions have a natural encoding under the belief function framework and can easily be modeled through the $CBNC$ representation.

In the last part of the thesis, we have proposed a model for causality ascription to identify influential relationships between different attributes of the system namely causality, facilitation or justification in the presence observational and interventional data. An intelligent artifact will look for the causes of an exceptional event that have changed the normal course of things to an abnormal situation. For that purpose, we have provided several definitions for the concepts of acceptance, rejection and ignorance. Since decision makers are not only interested in the presence of a causal link, our proposed model allows defining different strengths of a causal relationship.

Finally, we have implemented our belief causal network and showed that the belief graph augmentation and the belief graph mutilation are equivalent methods to deal with interventions. Besides, we have illustrated the feasibility and the usefulness of our belief causality ascription model through several scenarios. In fact, an agent will be able to discriminate between potential causes and ascribe the ones of the abnormal event. Through these scenarios, we have pointed out that our model can handle n-ary variables. Besides, we have shown that causes can be a disjunction of hypotheses and that effects can be represented as subsets of the cartesian product of some variables. Then, we have explained the usefulness of our model through a scenario in the context of the computer security area.

Some interesting future works have to be mentioned. First, we can explore belief causal inference and then extend the proposed tools for a complete mechanism of causal reasoning in presence of observational and interventional data. We can also study the relationship between interventions and belief changes in a belief function framework and analyze its relation with the belief function counterpart of Jeffrey's rule (Smets, 1993c) under uncertain inputs. We will investigate if this rule can encode the concept of interventions in belief causal graphical models and if it guarantees the uniqueness of the solution.

Another contribution regarding the causality ascription model is to develop an axiomatic system that includes formal properties of causation and facilitation. Moreover, we can take into account hidden causes in our model. Besides, we can investigate the counterpart of the Semi Markovian Causal Models (Tian & Pearl, 2002) which are an extension of causal Bayesian networks for modeling problems with latent variables. However, in theses models there is a problem of identifiability when making causal reasoning. In fact, we are not neces-

sarily able to identify the effects of an intervention on some variables. It will be interesting to study if it will be the same for their belief counterparts.

Another line or research will be to handle particular belief functions (e.g., qualitative, continuous belief function). Accordingly, we can compare the results of qualitative non-monotonic consequence model with the qualitative belief model.

From application point of view and in parallel to further theoretical elaboration, the belief causality ascription model can be used in the context of the computer network security system in the areas of intrusion detection. Indeed, we will develop methods allowing the ascription of several types of causes of attacks that can be either very strongly, strongly or weakly rejected events in the normal course of things. Note that our approach can also be used in other fields (e.g., marketing).

Appendix A

Proofs of Chapter 4

A.1 Introduction

In this appendix, we recall the different propositions given in Chapter 4 and provide their proofs.

A.2 Proofs of propositions

Proposition 4.1 Let $\mathcal{G}$ be a belief causal network and $m^V_{\mathcal{G}}$ be the joint mass distribution related to $\mathcal{G}$. The effect of an intervention $do(a_{ij})$ on the mass distribution is given by:

$$m^V_{\mathcal{G}}(v|do(a_{ij})) = \begin{cases} \sum\limits_{v' \cap a_{ij}^{\uparrow V} = v} m_{\mathcal{G}}^{V \setminus A_i}(v')^{\uparrow V} & \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

where v is a subset from the cartesian product of the variables in V.

Proof of Proposition 4.1 In Definition 4.1, we have explained that an intervention on the belief causal graph corresponds to an observation on the mutilated graph:

$m^V_{\mathcal{G}}(.|do(a_{ij})) \quad = m^V_{\mathcal{G}_{mut}}(.|a_{ij})$.

Using Dempster's rule of conditioning, we have :

(1) $m^V_{\mathcal{G}_{mut}}(.|a_{ij}) = m^V_{\mathcal{G}_{mut}} \oplus m^{A_i}_{\mathcal{G}_{mut}}$, with $m^{A_i}_{\mathcal{G}_{mut}}$ a certain *bba* focused on a_{ij}.

Note that using the independence principle, $m^V_{\mathcal{G}_{mut}}$ can be written as a combination of local distributions. In particular,

(2) $m^V_{\mathcal{G}_{mut}} = m^{V \setminus A_i \uparrow V}_{\mathcal{G}_{mut}} \oplus m^{A_i}_{\mathcal{G}_{mut}}(. \mid Pa(A_i))^{\uparrow V}$.

Intervening on A_i amounts to cutting-off the links relating it to its initials causes $PA(A_i)$. Thus, A_i becomes independent of $PA(A_i)$. Formally,

(3) $m^{A_i} = m^{A_i}(. \mid Pa(A_i))$.

On the other hand, $m^{A_i}_{\mathcal{G}_{mut}}$ is a certain *bba*. Then,

(4) $m^{A_i}_{\mathcal{G}_{mut}} \oplus m^{A_i}_{\mathcal{G}_{mut}} = m^{A_i}_{\mathcal{G}_{mut}}$.

Using (1), (2), (3) and (4), we obtain:

(5) $m^V_{\mathcal{G}_{mut}}(. \mid a_{ij}) = m^{V \setminus A_i \uparrow V}_{\mathcal{G}_{mut}} \oplus m^{A_i}_{\mathcal{G}_{mut}}$.

Dempster's rule of combination is defined as the orthogonal sum of two *bba*s, here $m_{\mathcal{G}_{mut}}^{V\setminus A_i\uparrow V}$ and $m_{\mathcal{G}_{mut}}^{A_i}$. Focal elements of the aggregated *bba* are all the possible intersections between pairs of focal elements of $m_{\mathcal{G}_{mut}}^{V\setminus A_i\uparrow V}$ and $m_{\mathcal{G}_{mut}}^{A_i}$.

$m_{\mathcal{G}_{mut}}^{A_i}$ is a certain *bba*. Thus, its focal element is a_{ij}. Let us denote by v' focal elements of $m_{\mathcal{G}_{mut}}^{V\setminus A_i\uparrow V}$ and by v elements of the aggregated *bba*. Hence, v will be considered as a focal element only if $v^{\downarrow A_i} = \{a_{ij}\}$. Thus, $m_{\mathcal{G}}^{V}(v|do(a_{ij}))$ will be defined as follows:

$$\begin{aligned} m_{\mathcal{G}}^{V}(v|do(a_{ij})) &= \sum_{v'\cap a_{ij}^{\uparrow V}=v} m_{\mathcal{G}}^{V\setminus A_i}(v')^{\uparrow V} \quad \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ &= 0 \qquad\qquad\qquad\qquad\quad \text{otherwise} \end{aligned}$$

Proposition 4.2 Let $\mathcal{G}$ be a belief causal network in which a variable A_i is forced it to take the value a_{ij}, $do(a_{ij})$. In the case where the initial *bba* of the target variable is non-dogmatic, the effect of this intervention on the joint mass distribution $m_{\mathcal{G}}^{V}$ is given as follows:

$$m_{\mathcal{G}}^{V}(v|do(a_{ij})) = \begin{cases} \sum\limits_{\mathcal{F}\cap a_{ij}^{\uparrow V}=v} m_{\mathcal{G}}^{V} \ominus (\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_j\times PA(A_j)})^{\uparrow V} & \text{if } v^{\downarrow A_i} = \{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

where $\mathcal{F}$ represent focal elements of $(m_{\mathcal{G}}^{V} \ominus (\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_j\times PA(A_j)})^{\uparrow V})$ and v any subset of the cartesian product of variables in V.

Proof of Proposition 4.2 From Proposition 4.1, two cases arise:

- If $v^{\downarrow A_i} = \{a_{ij}\}$, then:

$m_{\mathcal{G}}^{V}(v|do(a_{ij})) = \sum\limits_{v'\cap a_{ij}^{\uparrow V}=v} m_{\mathcal{G}}^{V\setminus A_i}(v')^{\uparrow V}$

On the other hand, using the independence principle, we have:

$m_{\mathcal{G}}^{V} = m_{\mathcal{G}}^{V\setminus A_i\uparrow V} \oplus (\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_i\times PA(A_i)})^{\uparrow V}$

Thus, when the *bba* of the target variable is non-dogmatic we can use the decombination operator as follows:

$m_{\mathcal{G}}^{V\setminus A_i\uparrow V} = m_{\mathcal{G}}^{V} \ominus (\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_j\times PA(A_j)})^{\uparrow V}$

Let us denote by $\mathcal{F}$, the set of focal elements of $(m_{\mathcal{G}}^{V}\ominus(\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_j\times PA(A_j)})^{\uparrow V})$. Thus,

$m_{\mathcal{G}}^{V}(v|do(a_{ij})) = \sum\limits_{\mathcal{F}\cap a_{ij}^{\uparrow V}=v} m_{\mathcal{G}}^{V} \ominus (\bigoplus\limits_{Pa(A_j)} m_{\mathcal{G}}^{A_i}(.|Pa(A_i))^{\Cap A_j\times PA(A_j)})^{\uparrow V}$

- If $v^{\downarrow A_i} \neq \{a_{ij}\}$, then the proof is immediate from Proposition 4.1. Clearly $m_{\mathcal{G}}^{V}(v|do(a_{ij})) = 0$

Corollary 4.1 Let $\mathcal{G}$ be a belief causal network whose joint mass distribution is $m_{\mathcal{G}}^{V}$. In the case where the initial *bba* of the target variable is non-dogmatic, the effect of an intervention $do(a_{ij})$ on a variable A_i of this graph forcing it to take the value a_{ij} can be also computed as

follows:

$$m^V_{\mathcal{G}}(v|do(a_{ij})) = \begin{cases} \sum\limits_{\mathcal{F}\cap\{(a_{ij},Pa(A_i))\}^{\uparrow V}=v} m^V_{\mathcal{G}}(.|a_{ij},Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V} & \text{if } v^{\downarrow A_i}=\{a_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

where $\mathcal{F}$ is a focal element of $m^V_{\mathcal{G}}(.|a_{ij},Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V}$ and v is a subset from the cartesian product of variables in V.

Proof of Corollary 4.1 From Proposition 4.2, we have shown that when the *bba* of the target variable is non-dogmatic:

- If $v^{\downarrow A_i}=\{a_{ij}\}$, then:

(1) $m^V_{\mathcal{G}}(v|do(a_{ij})) = \sum\limits_{\mathcal{F}\cap a_{ij}^{\uparrow V}=v} m^V_{\mathcal{G}} \ominus (\bigoplus\limits_{Pa(A_j)} m^{A_i}_{\mathcal{G}}(.|Pa(A_i))^{\wr A_j\times PA(A_j)})^{\uparrow V}$

Using conditioning, the *bba* $m^V_{\mathcal{G}}$ is computed as:

(2) $m^V_{\mathcal{G}}(.|a_{ij},Pa(A_i)) \oplus m^{A_i\times PA(A_i)}_{\mathcal{G}}(\{(a_{ij},Pa(A_i))\})^{\uparrow V}$

Where,

(3) $m^{A_i\times PA(A_i)}_{\mathcal{G}}(\{(a_{ij},Pa(A_i))\})^{\uparrow V}) = m^{A_i}_{\mathcal{G}}(.|Pa(A_i))^{\wr A_j\times PA(A_j)})^{\uparrow V} \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V}$

From (1), (2) and (3), $m^V_{\mathcal{G}}(v|do(a_{ij}))$ is given by:

$$\sum_{\mathcal{F}\cap\{(a_{ij},Pa(A_i))\}^{\uparrow V}=v} m^V_{\mathcal{G}}(.|a_{ij},Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V}$$

where $\mathcal{F}$ is a focal element of $m^V_{\mathcal{G}}(.|a_{ij},Pa(A_i)) \oplus m^{PA(A_i)}_{\mathcal{G}}(Pa(A_i))^{\uparrow V}$

- If $v^{\downarrow A_i}=\{a_{ij}\}$, then $m^V_{\mathcal{G}}(v|do(a_{ij}))=0$.

Proposition 4.3 Let $\mathcal{G}_{aug}$ be an augmented causal belief graph where the DO node is set to the value nothing.
Its corresponding *bba* $m^V_{\mathcal{G}_{aug}}(.|do(nothing))$ encodes the same joint distribution as the original causal belief graph.

$$m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^V_{\mathcal{G}} \tag{A.4}$$

where $V' = V \cup DO$.

Proof of Proposition 4.3 Let us denote by V' the set of the variables in the graph, i.e., $V' = V \cup DO$. The global mass distribution on the mutilated graph $m^{V'}_{\mathcal{G}_{aug}}$ is computed as:

$$\begin{aligned} m^{V'}_{\mathcal{G}_{aug}} &= m^{V\wr V'}_{\mathcal{G}_{aug}} \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}} \\ &= m^{V\setminus A_i\wr V'}_{\mathcal{G}_{aug}} \oplus m^{A_i\wr V'}_{\mathcal{G}_{aug}}(.|Pa'(A_i)) \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}} \end{aligned}$$

Conditioning $m^{V'}_{\mathcal{G}_{aug}}$ on $do(nothing)$, corresponds to observing the value $nothing$ on the network:

$m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^{V\setminus A_i\wr V'}_{\mathcal{G}_{aug}} \oplus m^{A_i}_{\mathcal{G}_{aug}}(.|Pa'(A_i))^{\wr V'} \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}} \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}}$.

In the case where m^{DO} is a certain *bba*, focused on $do(nothing)$, m^{DO} is idempotent, i.e., $m^{DO} \oplus m^{DO} = m^{DO}$.

Therefore, using Equation 4.6 $m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing))$ becomes:

$$\begin{aligned} &= m^{V\setminus A_i\wr V'}_{\mathcal{G}_{aug}} \oplus m^{A_i}_{\mathcal{G}_{aug}}(.|Pa(A_i))^{\wr V'} \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}} \\ &= m^{V\wr V'}_{\mathcal{G}} \oplus m^{DO\wr V'}_{\mathcal{G}_{aug}} \end{aligned}$$

Using Equation 4.13, we get: $m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^{V}_{\mathcal{G}}$.

Proposition 4.4 Let $\mathcal{G}$ be a causal belief network and $\mathcal{G}_{mut}$ and $\mathcal{G}_{aug}$ its corresponding mutilated and augmented graphs after acting on the variable A_i by forcing it to take the value a_{ij}. Then as for probability and possibility theories, computing the effects of interventions using the mutilation of the graph or its augmentation gives the same results.

$$\begin{array}{ll} & m^{V}_{\mathcal{G}}(.|do(a_{ij})) \\ = & m^{V}_{\mathcal{G}_{mut}}(.|a_{ij}) \\ = & m^{V}_{\mathcal{G}_{aug}}(.|do(a_{ij})) \end{array}$$

Proof of Proposition 4.4 Let $\mathcal{G}_{aug}$ be an augmented causal belief network on which an intervention is performed on a variable A_i setting it to the value a_{ij}, i.e., $do(x) = do(a_{ij})$. From Equation 4.1, we have $m^{V}_{\mathcal{G}}(.|do(a_{ij})) = m^{V}_{\mathcal{G}_{mut}}(.|a_{ij})$.

Let us denote by V' the set of the variables in the graph, i.e., $V' = V \cup DO$. According to the conditional independence principle, $m^{V'}_{\mathcal{G}_{aug}}$ is given by:

$m^{V\setminus A_i \Uparrow V'}_{\mathcal{G}_{aug}} \oplus m^{A_i}_{\mathcal{G}_{aug}}(.|Pa'(A_i))^{\Uparrow V'} \oplus m^{DO \Uparrow V'}_{\mathcal{G}_{aug}}$.

When m^{DO} is a certain *bba* focused on a_{ij}, from Equation 4.6 and Equation 4.9 we get: $m^{V}_{\mathcal{G}_{aug}}(.|do(a_{ij})) = m^{V}_{\mathcal{G}}(.|do(a_{ij}))$.

Accordingly, $m^{V}_{\mathcal{G}_{aug}}(.|do(a_{ij})) = m^{V}_{\mathcal{G}}(.|do(a_{ij})) = m^{V}_{\mathcal{G}_{mut}}(.|a_{ij})$

Proposition 4.5 The global joint distribution obtained after intervening on the initial graph $\mathcal{G}$ by forcing a variable A_i to take the imprecise value sub_{ik} leads to the same results obtained after observing sub_{ik} on the mutilated graph $\mathcal{G}_{mut}$:

$$m^{V}_{\mathcal{G}_{mut}}(.|sub_{ik}) = m^{V}_{\mathcal{G}}(.|do(sub_{ik}))$$

Proof of Proposition 4.5 Same as the proof presented for standard interventions, i.e., proof of Proposition 4.1.

Proposition 4.6 Let $\mathcal{G}$ a belief causal network and let $\mathcal{G}_{mut}$ and $\mathcal{G}_{aug}$ its corresponding mutilated and augmented graphs. Dealing with imprecise interventions using the mutilation of the graph or its augmentation gives the same results.

$$\begin{array}{ll} & m^{V}_{\mathcal{G}}(.|do(sub_{ik})) \\ = & m^{V}_{\mathcal{G}_{mut}}(.|sub_{ik}) \\ = & m^{V}_{\mathcal{G}_{aug}}(.|do(sub_{ik})) \end{array}$$

Proof of Proposition 4.6 The proof presented for standard interventions remains valid here.

Proposition 4.7 Standard interventions are a particular case of imprecise interventions when the the subset representing the possible target values is composed of one element, i.e. $sub_{ij} = \{a_{ij}\}$.

$$m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik} = \{a_{ij}\}, sub_{ik} \subseteq \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases}$$

Proposition 4.8 Interventions with ignored target values are a particular case of imprecise interventions: the subset representing the possible target values corresponds to the frame of discernment, i.e. $sub_{ij} = \Theta_{A_i}$.

$$m^{DO}(do(x)) = \begin{cases} 1 & \text{if } x = sub_{ik}, sub_{ik} = \Theta_{A_i} \\ 0 & \text{otherwise} \end{cases}$$

Proofs of Proposition 4.7 and Proposition 4.8 Proofs are immediate. The belief function theory allows to allocate beliefs to subsets. A subset includes elementary events and the whole frame of discernment of a given variable.

Proposition 4.9 Standard interventions are a particular case of uncertain interventions when the source is fully reliable, i.e., $\alpha = 0$.

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} 1 & \text{if } sub_{ik} = a_{ij} \\ 0 & \text{otherwise} \end{cases}$$

Proof of Proposition 4.9 The proof is immediate by the application of Equation 4.21.

Proposition 4.10 The beliefs provided about the non-occurrence of an intervention are accepted without any modification. They are defined like standard interventions.

$$m^{A_i,\alpha_{do(nothing)}=0}(.|do(nothing)) = m^{A_i}(.|do(nothing))$$

Proof of Proposition 4.10 The non-occurrence of the intervention is a certain fact. Thus $\alpha_{do(nothing)} = 0$, amounts to keep the same mass distribution defined in Equation 4.26.

Proposition 4.11 An augmented belief function causal graph where the DO node is set to the value nothing encodes the same joint distribution than the initial causal belief network.

$$m^{V'}_{\mathcal{G}_{aug}}(.|do(nothing)) = m^{V}_{\mathcal{G}}$$

where $V' = V \cup DO$.

Proof of Proposition 4.11 The proof provided for Proposition 4.3 remains valid for Proposition 4.11.

Proposition 4.12 Uncertain interventions with a certain consequence are a particular case of uncertain ones with uncertain consequences when the parameter β_j is set to one.

$$m^{A_i}(sub_{ik}|do(a_{ij})) = \begin{cases} 1-\alpha & \text{if } sub_{ik} = a_{ij} \\ \alpha & \text{if } sub_{ik} = \Theta_{A_i} \end{cases}$$

Proof of Proposition 4.12 A certain consequence means that it exists one focal element which is the target value a_{ij}. The proof is immediate by the application of the Equation 4.31 for $\beta_j = 1$.

A.3 Conclusion

In this appendix, we have given proofs of the different propositions that we have provided in Chapter 4.

Appendix B

Proofs of Chapter 5

B.1 Introduction

In this appendix, we recall the different propositions given in Chapter 5 and provide their proofs.

B.2 Proofs of propositions

Proposition 5.1 If $\Theta_E = \{e_i, \overline{e_i}\}$ then:

- e_i is weakly accepted iff e_i is strongly accepted iff e_i is very strongly accepted iff $\text{pl}(e_i) > \text{pl}(\overline{e_i})$.
- e_i is weakly rejected iff e_i is strongly rejected iff e_i is very strongly rejected if $\text{pl}(e_i) < \text{pl}(\overline{e_i})$.

Proof of Proposition 5.1 All the definitions presented in Definition 5.1 collapsed in the case where $\Theta_E = \{e_i, \overline{e_i}\}$.

Proposition 5.2 If an event e_i is very strongly rejected in a given context and *after* observing an event e_j it becomes very strongly accepted, then e_j is said to be a *very strong cause* of e_i, namely

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c);$$

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$$

Proof of Proposition 5.2 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a very strong reject. Thus, $pl^t(e_i|c) < pl^t(\overline{e_i}|c)$. After the observation of the event e_j, e_i becomes very strongly accepted. To compute the effect of observations, we have to use conditioning. Accordingly, we get: $pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$.

Proposition 5.3 If an event e_i is very strongly rejected in a given context and *after* observing an event e_j it becomes strongly accepted, then e_j is said to be a *strong cause* of e_i, namely

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c);$$

$\exists e \neq e_i$,

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proof of Proposition 5.3 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a very strong reject. Thus, $pl^t(e_i|c) < pl^t(\overline{e_i}|c)$. After the observation of the event e_j, e_i becomes strongly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.4 If an event e_i is strongly rejected in a given context and *after* observing an event e_j it becomes very strongly accepted, then e_j is said to be a *strong cause* of e_i, namely

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$$

Proof of Proposition 5.4 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a strong reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

After the observation of the event e_j, e_i becomes very strongly accepted. To compute the effect of observations, we have to use conditioning. Accordingly, we have: $pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$.

Proposition 5.5 If an event e_i is very strongly rejected in a given context and *after* observing an event e_j it becomes weakly accepted, then e_j is said to be a *weak cause* of e_i, namely

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c)$$

$\exists e \neq e_i$,

$$pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c)) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

Proof of Proposition 5.5 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a very strong reject. Thus, $pl^t(e_i|c) < pl^t(\overline{e_i}|c)$. After the observation of the event e_j, e_i becomes weakly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c)) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

Proposition 5.6 If an event e_i is strongly rejected in a given context and *after* observing an event e_j it becomes strongly accepted, then e_j is said to be a *weak cause* of e_i, namely

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

$\exists e \neq e_i$,

$$pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c)) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

Proof of Proposition 5.6 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a strong reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

After the observation of the event e_j, e_i becomes strongly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c)) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

Proposition 5.7 If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes very strongly accepted, then e_j is said to be *a weak cause* of e_i namely,

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c)$$
$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$$

Proof of Proposition 5.7 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a weak reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

After the observation of the event e_j, e_i becomes very strongly accepted. To compute the effect of observations, we have to use conditioning. Accordingly, we have: $pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$.

Proposition 5.8 If an event e_i is strongly rejected in a given context and after observing an event e_j it becomes weakly accepted, then e_j is said to be a very weak cause of e_i, namely

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$
$$\text{where } pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

Proof of Proposition 5.8 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a strong reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

After the observation of the event e_j, e_i becomes weakly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.9 If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes strongly accepted, then e_j is said to be a very weak cause of e_i, namely

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c)$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$
$$\text{where } \; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

Proof of Proposition 5.9 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a weak reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

After the observation of the event e_j, e_i becomes strongly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

An event is strongly accepted if it is not very strongly accepted. Accordingly, (2)

$$pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j,c) < pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c)) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$

Proposition 5.10 If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes weakly accepted, then e_j is said to be a slight cause of e_i, namely

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c)$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j,c) \leq pl^{t+n}(e_i|e_j,c) \leq pl^{t+n}(\overline{e_i}|e_j,c)$$
$$\text{where } \; pl^{t+n}(e_i|e_j,c) = Argmax(pl^{t+n}(e_k|e_j,c))$$

Proof of Proposition 5.10 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a weak reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

After the observation of the event e_j, e_i becomes weakly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.11 to Proposition 5.19 (see Section 5.3.3)

Proofs of Proposition 5.11 to Proposition 5.19 The proofs of propositions concerning the causality ascription in presence of interventional data are similar to the ones in presence of observational data. However, instead of conditioning we have to apply causal conditioning using the do operator under the belief function framework as it has been explained in Chapter 4.

Proposition 5.20 An event e_j is said to confirm another event e_i if:

$$pl^t(e_i) \cdot pl^{t+n}(e_j) < pl^{t+n}(e_i, e_j)$$

Proof of Proposition 5.20 According to Definition 1.2 we have,

$$pl(e_i|e_j) > pl(e_i)$$

Using conditioning, we have:

$$\frac{pl^{t+n}(e_i, e_j)}{pl^{t+n}(e_j)} > pl^t(e_i)$$

It is equivalent to:

$$pl^{t+n}(e_i, e_j) > pl^t(e_i) \cdot pl^{t+n}(e_j)$$

Proposition 5.21 An event e_j is said to attenuate another event e_i if:

$$pl(e_i, e_j) < pl(e_i) \cdot pl(e_j)$$

Proof of Proposition 5.21 According to Definition 1.3 we have,

$$pl(e_i|e_j) < pl(e_i)$$

. Using conditioning, we have:

$$\frac{pl^{t+n}(e_i, e_j)}{pl^{t+n}(e_j)} < pl^t(e_i)$$

It is equivalent to:

$$pl^{t+n}(e_i, e_j) < pl^t(e_i) \cdot pl^{t+n}(e_j)$$

Proposition 5.22 If an event e_i is very strongly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to very strongly facilitate the occurrence of e_i. Namely,

$$pl^t(e_i|c) < pl^t(\overline{e_i}|c).$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proof of Proposition 5.22 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a very strong reject. Thus, $pl^t(e_i|c) < pl^t(\overline{e_i}|c)$. After the observation of the event e_j, e_i becomes ignored. Accordingly, we obtain:

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proposition 5.23 If an event e_i is strongly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to strongly facilitate the occurrence of e_i. Namely,

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proof of Proposition 5.23 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a strong reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) < pl^t(e|c)$$

After the observation of the event e_j, e_i becomes ignored. Accordingly, we obtain:

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proposition 5.24 If an event e_i is weakly rejected in a given context and after observing an event e_j it becomes ignored then e_j is said to weakly facilitate the occurrence of e_i. Namely,

$$\exists e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_j|c)) \leq pl^t(e|c)$$

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proof of Proposition 5.24 The effect e_i is abnormal in context c. The definition of abnormality is here defined as a weak reject. Thus, (1)

$$pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

Since very strongly rejected is not verified, (2)

$$pl^t(\overline{e_i}|c) \leq pl^t(e_i|c)$$

From (1) and (2), we have

$$\forall e \neq e_i, pl^t(\overline{e_i}|c) \leq pl^t(e_i|c) = Argmin(pl^t(e_k|c)) \leq pl^t(e|c)$$

After the observation of the event e_j, e_i becomes ignored. Accordingly, we obtain:

$$pl^{t+n}(e_i|e_j, c) = pl^{t+n}(\overline{e_i}|e_j, c).$$

Proposition 5.25 If an event e_i is ignored in a given context and after observing an event e_j it becomes very strongly accepted then e_j is said to very strongly justify the occurrence of e_i. Namely,

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c)$$

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$$

Proof of Proposition 5.25 The event e_i is ignored in a given context. Accordingly, we have:

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c)$$

After the observation of the event e_j, e_i becomes very strongly accepted. To compute the effect of observations, we have to use conditioning. Accordingly, we get:

$$pl^{t+n}(e_i|e_j, c) > pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.26 If an event e_i is ignored in a given context and after observing an event e_j it becomes strongly accepted then e_j is said to strongly justify the occurrence of e_i. Namely,

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c)$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$
$$\text{where } \; pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

Proof of Proposition 5.26 The event e_i is ignored in a given context. Accordingly, we have:

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c)$$

After the observation of the event e_j, e_i becomes strongly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j, c) < pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.27 If an event e_i is ignored in a given context and after observing an event e_j it becomes weakly accepted then e_j is said to weakly justify the occurrence of e_i. Namely,

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c).$$

$$\exists e \neq e_i, pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$
$$\text{where } \; pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

Proof of Proposition 5.27 The event e_i is ignored in a given context. Accordingly, we have:

$$pl^t(e_i|c) = pl^t(\overline{e_i}|c)$$

After the observation of the event e_j, e_i becomes weakly accepted. To compute the effect of observations, we have to use conditioning. Using the definition of conditional strong accept, we get: (1)

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c))$$

An event is strongly accepted if it is not very strongly accepted, Accordingly, (2)

$$pl^{t+n}(e_i|e_j, c) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

From (1) and (2), we have:

$$pl^{t+n}(e|e_j, c) \leq pl^{t+n}(e_i|e_j, c) = Argmax(pl^{t+n}(e_k|e_j, c)) \leq pl^{t+n}(\overline{e_i}|e_j, c)$$

Proposition 5.28 to Proposition 5.33 (see Section 5.6)

Proofs of Proposition 5.28 to Proposition 5.33 As for causality ascription, the proofs of propositions concerning the facilitation and justification ascription in presence of interventional date are similar to the ones in presence of observational data. However, instead of conditioning we have to apply causal conditioning using the do operator under the belief function framework as it has been explained in Chapter 4.

B.3 Conclusion

In this appendix, we have provided proofs of the different propositions that we have provided in Chapter 5.

References

Adams, E. (1975). *The logic of conditionals.* D. Reidel Publishing Co.

Appriou, A. (1999). Multisensor signal processing in the framework of the theory of evidence. In *Application of Mathematical Signal Processing Techniques to Mission Systems* (Vol. 216 Lecture Series).

Barnett, J. A. (1981). Computational methods for a mathematical theory of evidence. In *International Joint Conference on Artificial Intelligence (IJCAI'81)* (pp. 868–875).

Ben Abdallah, N., Mouhous Voyneau, N., & Denoeux, T. (2012). Combining statistical and expert evidence within the d-s framework: Application to hydrological return level estimation. In *International Conference on Belief Functions (BELIEF'12)* (Vol. 164 ASC, p. 393-400). Springer-Verlag.

Ben Amor, N., Benferhat, S., & Mellouli, K. (2003). Anytime propagation algorithm for min-based possibilistic graphs. *Soft Computing*, *8*(2), 150-161.

Ben Yaghlane, B., & Mellouli, K. (2008). Inference in directed evidential networks based on the transferable belief model. *International Journal of Approximate Reasoning*, *48*(399-418).

Ben Yaghlane, B., Smets, P., & Mellouli, K. (2002a). Belief function independence: II. the conditional case. *International Journal of Approximate Reasoning*, *31*, 31–75.

Ben Yaghlane, B., Smets, P., & Mellouli, K. (2002b). Belief function independence: I. the marginal case. *International Journal of Approximate Reasoning*, *29*, 47–70.

Ben Yaghlane, B., Smets, P., & Mellouli, K. (2003). Directed evidential networks with conditional belief functions. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03)* (Vol. 2711 LNCS). Springer-Verlag.

Benavoli, A., Ristic, B., Farina, A., Oxenham, M., & Chisci, L. (2009). An application of evidential networks to threat assessment. *IEEE Transactions on Aerospace and Electronic Systems*, *45*, 620-639.

Benferhat, S. (2010). Interventions and belief change in possibilistic graphical models. *Artificial Intelligence*, *174*(2), 177-189.

Benferhat, S., Bonnefon, J., Chassy, P., Da Silva Neves, R., Dubois, D., Dupin de Saint Cyr-Bannay, F., et al. (2008). A comparative study of six formal models of causal ascription. In *International Conference on Scalable Uncertainty Management (SUM'08)* (Vol. 5291 LNCS, pp. 47–62). Springer - Verlag.

Benferhat, S., Dubois, D., & Prade, H. (1992). Representing default rules in possibilistic logic. In (p. 673-684).

Benferhat, S., Saffiotti, A., & Smets, P. (2000). Belief functions and default reasoning. *Artificial Intelligence*, *122*(1-2), 1-69.

Benferhat, S., & Smaoui, S. (2007a). Hybrid possibilistic networks. *International Journal of Approximate Reasoning*, *44*(3), 224-243.

Benferhat, S., & Smaoui, S. (2007b). Possibilistic causal networks for handling interventions: A new propagation algorithm. In *AAAI Conference on Artificial Intelligence (AAAI'07)* (p. 373-378). AAAI Press.

Benferhat, S., & Smaoui, S. (2008). Quantitative possibilistic networks: Handling interventions. In *Mexican International Conference on Artificial Intelligence (MICAI'08)* (Vol. 5317 LNCS, pp. 720–731). Springer-Verlag.

Benferhat, S., & Smaoui, S. (2011). Inferring interventions in product-based possibilistic causal networks. *Fuzzy Sets and Systems*, *169*(1), 26-50.

Bloch, I. (1996). Information combination operator for data fusion: a comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*, 52–67.

Bonnefon, J., Da Silva Neves, R., Dubois, D., & Prade, H. (2006). Background default knowledge and causality ascriptions. In *European conference on artificial intelligence (ecai'06)* (Vol. 141, p. 11-15). IOS Press.

Bonnefon, J., Da Silva Neves, R., Dubois, D., & Prade, H. (2008). Predicting causality ascriptions from background knowledge: model and experimental validation. *International Journal of Approximate Reasoning*, *48*(3), 752-765.

Boukhris, I., Benferhat, S., & Elouedi, Z. (2011a). A belief function model for ascribing causality. In *Portuguese Conference on Artificial Intelligence (EPIA'11)* (p. 342-356).

Boukhris, I., Benferhat, S., & Elouedi, Z. (2011b). Representing belief function knowledge with directed acyclic graphs. In *International Conference on Knowledge Science, Engineering and Management (KSEM'11)* (Vol. 7091 LNAI, p. 233-245). Springer-Verlag.

Boukhris, I., Benferhat, S., & Elouedi, Z. (2012a). Ascribing causality from interventional belief function knowledge. In *International Conference on Belief Functions (BELIEF'12)* (Vol. 164 ASC, p. 229-237). Springer-Verlag.

Boukhris, I., Benferhat, S., & Elouedi, Z. (2012b). Ascribing strong causes from observational and interventional data under the belief function framework. In *International FLINS Conference on Uncertainty Modeling in Knowledge Engineering and Decision Making (FLINS'12).* World Scientific.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2011a). Analyzing belief networks with conditional beliefs. In *International Conference on Intelligent Systems Design and Applications (ISDA'11)* (p. 959-964). IEEE Computer Society.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2011b). Modeling interventions using belief causal networks. In *International Florida Artificial Intelligence Research Society Conference (FLAIRS'11)* (p. 602-607). AAAI Press.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2012a). Acceptation relations to causality ascription in a belief function framework. In *International Symposium on Artificial Intelligence and Mathematics (ISAIM'12).*

Boukhris, I., Elouedi, Z., & Benferhat, S. (2012b). Dealing with interventions with uncertain consequences in belief causal networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'12)* (Vol. 299 CCIS, p. 585-595). Springer-Verlag.

Cano, J., Delgado, M., & Moral, S. (1993). An axiomatic framework for propagating uncer-

tainty in directed acyclic networks. *International Journal of Approximate Reasoning*, *8*, 253–280.

Chassy, P., Dubois, D., & Prade, H. (2011). Understanding what is going on, or how to make sense of situations. In *International Workshop on Uncertainty Reasoning and Multi-Agent Systems for Sensor Networks.*

Cooke, R., & Smets, P. (2001). Self-conditional probabilities and probabilistic interpretations of belief functions. *Annals of Mathematics and Artificial Intelligence*, *32*(1-4), 269-285.

Cozman, F. (1997). Robustness analysis of Bayesian networks with local convex sets of distributions. In *Uncertaintyin Artificial Intelligence (UAI'97)* (p. 108-115).

Cozman, F. (2000). Credal networks. *Artificial Intelligence*, *120*, 199-233.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks* (Hardcover, Ed.). Cambridge University Press.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, *38*, 325–339.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, *30*, 205–247.

Denœux, T., & Zouhal, L. M. (2001). Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, *122*, 47–62.

Dickerson, J., & Dickerson, J. (2000). Fuzzy network profiling for intrusion detection. In *International Conference of the North American Fuzzy Information Processing Society (NAFIPS'00)* (pp. 301–306).

Dubois, D., & Denœux, T. (2010). Statistical inference with belief functions and possibility measures: A discussion of basic assumptions. In *Combining Soft Computing and Statistical Methods in Data Analysis* (Vol. 77 ASC, p. 217-225). Springer-Verlag.

Dubois, D., Grabisch, M., Prade, H., & Smets, P. (2001). Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: the assessment of the value of a candidate. *International Journal of Intelligent Systems*, *16*, 1245–1272.

Dubois, D., & Prade, H. (198). The principle of minimum specificity as a basis for evidential reasoning. In *International Conference on Informtion Processing and Management of Uncertainty in Knowledge-Based System (IPMU'86)* (Vol. 286 LNCS, pp. 75–84). Springer-Verlag.

Dubois, D., & Prade, H. (1982). On several representations of an uncertainty body of evidence. In *Fuzzy Information and Decision Processes* (p. 167-181).

Dubois, D., & Prade, H. (1986). On the unicity of Dempster rule of combination. *International Journal of Intelligent System*, *1*, 133-142.

Dubois, D., & Prade, H. (1988). *Possibility theory : an approach to computerized processing of uncertainty.* Plenum Press.

Dubois, D., & Prade, H. (1998). Possibility theory: qualitative and quantitative aspects. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (Vol. 1, pp. 169–226).

Dubois, D., & Prade, H. (2003). Liens causaux et explications. problèmes de modélisation: une discussion préliminaire. *Journées Nationales sur les Modèles de Raisonnement GDR I3 Information-Interaction-Intelligence*, 81-90.

Dubois, D., Prade, H., Bonnefon, J., & Da Silva Neves, R. (2009). *Causal ascription based on context-dependent subjective knowledge.* Workshop on Modèles Informatiques et Cognitifs du Raisonnement Causal.

Duncan, O. (1975). *Introduction to structural equation models.* New york: Academic Press.

Dupin de Saint-Cyr, F. (2008). Scenario update applied to causal reasoning. In *International Conference on Principles and Knowledge Representation and Reasoning (KR'08)* (p. 188-197).

Dupin de Saint-Cyr, F., & Lang, J. (2002). Belief extrapolation (or how to reason about observations and unpredicted change). In *International Conference on Principles and Knowledge Representation and Reasoning (KR'02)* (p. 497-508).

Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. In *Philosophy of Science Association* (Vol. 74, p. 981-995).

Eells, E. (1991). *Probabilistic causality.* Cambridge: Cambridge University Press.

Elouedi, Z., Mellouli, K., & Smets, P. (2001). Belief decision trees: theoretical foundations. *International Journal of Approximate Reasoning*, *28*(2-3), 91-124.

Fonck, P. (1994). Conditional independence in possibility theory. In *Uncertainty in Artificial Intelligence (UAI'94)* (pp. 221–226).

Foucher, S., Laliberte, F., Boulianne, G., & Gagnon, L. (2006). A Dempster-Shafer based fusion approach for audio-visual speech recognition with application to large vocabulary french speech. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP'06)* (Vol. 1, p. 597-600).

Gänderfors, P. (1992). *Belief revision.* Cambridge University Press.

Giunchiglia, E., Lee, J., McCain, N., Lifschitz, V., & Turner, H. (2004). Nonmonotoni causal theories. *Artificial Intelligence*, *153*, 49-104.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology.* Cambridge: MIT Press.

Goldszmidt, M., & Pearl, J. (1992). Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In *International Conference on Principles of Knowledge Representation and Reasoning (KR'92)* (p. 661-672).

Good, I. (1961a). A causal calculus I. *British Journal for the Philosophy of Science*, *11*, 305-318.

Good, I. (1961b). A causal calculus II. *British Journal for the Philosophy of Science*, *12*, 43-51.

Gordon, J., & Shortliffe, E. H. (1984). The Dempster-Shafer theory of evidence. In *Rule-Based Expert Systems: the MYCIN Experiments of the Stanford Heuristic Programming Project* (pp. 272–292).

Granger, C. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, *2*, 329-352.

Guan, J. W., & Bell, D. A. (1991). *Evidence theory and its applications.* Elsevier science.

Halpern, J., & Pearl, J. (2005). Causes and explanations: A structurel model approach. Part i: Causes. *British Journal for the Philosophy of Science*, *56*(45), 843-887.

Hart, H., & Honoré, T. (1985). *Causaltion in the law.* Oxford University Press.

Hilton, D., & Slugoski, B. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75-88.

Hsia, Y. T. (1991). Characterizing belief with minimum commitment. In *International Joint Conference on Artificial Intelligence (IJCAI'91)* (pp. 1184–1189).

Hume, D. (2006). *An enquiry concerning human understanding.* Digireads.com.

Jensen, F., Lauritzen, S., & Olesen, K. (1990). Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, *4*, 269-282.

Jensen, F., & Nielsen, T. (2007). *Bayesian networks and decision graphs.* Springer.

Katsuno, H., & Mendelzon, A. O. (1991). On the difference between updating a knowledge

base and revising it. In *International Conference on Principle of Knowledge Representation and Reasoning (KR'91)* (pp. 87–394).

Kayser, D., & Mokhtari, A. (1998). Times in a causal theory. *Annals of Mathematics and Artificial Intelligence*, 117-138.

Kayser, D., & Nouioua, F. (2005). About norms and causes. *International Journal on Artificial Intelligence Tools*, 7-23.

Khelfallah, M., & Mokhtari, A. (2001). Ramification in the normative method of causality. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01)* (Vol. 2143 LNCS, p. 704-713). Springer-Verlag.

Kim, H., & Swain, P. H. (1995). Evidential reasoning approach to multisource-data classification in remote sensing. *IEEE Transactions on Systems, Man and Cybernetics*, *25*(8), 1257–1265.

Klir, G., & Wierman, M. J. (1998). *Uncertainty-based-information : elements of generalized information theory.* Physica-Verlag.

Kohlas, J., & Monney, P. (1995). A mathematical theory of hints: An approach to Dempster-Shafer theory of evidence. In Springer-Verlag (Ed.), *Lecture Notes in Economics and Mathematical Systems.*

Korb, K., Hope, L., Nicholson, A., & Axnick, K. (2004). Varieties of causal intervention. In *Pacific Rim International Conference on Artificial Intelligence (PRICAI'04)* (Vol. 3157 LNCS, p. 322-331). Springer-Verlag.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*, 167-207.

Laâmari, W., Ben Yaghlane, B., & Simon, C. (2010). Comparing evidential graphical models for imprecise reliability. In *International Conference Scalable Uncertainty Management (SUM'10)* (Vol. 6379 LNCS, p. 191-204).

Lefevre, E., Vannoorenberghe, P., & Colot, O. (2000). Evience theory and color image segmentation. In *International Conference on Color in Graphics and Image Processing (ICCGIP'00)* (p. 164-169).

Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, *55*, 1-60.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297–315.

Lewis, D. (2004). Causation as influence. In (p. 75-106). Collins, Hall, and Paul.

Meganck, S., Leray, P., & Manderick, B. (2006). Learning causal Bayesian networks from observations and experiments: A decision theoretic approach. In *International Conference on Modeling Decisions for Artificial Intelligence (MDAI'06)* (Vol. 3885 LNCS, p. 58-69). Springer-Verlag.

Mellor, D. (1995). *The facts of causation.*

Otero, M., & Otero, R. (2000). Using causality for diagnosis. In *International Workshop on Principles of Diagnosis (DX'00)* (pp. 171–176).

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Pub. San Mateo, Ca, USA.

Pearl, J. (1998). Graphical models for probabilistic and causal reasonning. In s. D.M Gabbay P.Smets (Ed.), (Vol. The Handbook of Defeasible Resonning and Uncertainty Management Systems, p. 367-389). Kluwer Academic Press.

Pearl, J. (2000). *Causality: Models, reasonning and inference.* Cambridge University Press.

Pearl, J., & Hopkins, M. (2007). Causality and counterfactuals in the situation calculus. *Journal of Logic and Computation*, *17*, 939-953.

Ristic, B., & Smets, P. (2005). Target identification using belief functions and implication rules. *IEEE Transactions on Aerospace and Electronic Systems*, *41*(3), 1097-1103.

Salmon, W. (1980). Probabilistic causality. In (p. 208-232). Oxford University Press.

Scheines, R. (2008). Causation, statistics and the law. *Journal of Law and Policy*, *16*.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton Univ. Press. Princeton, NJ.

Shafer, G. (1982). Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, *44*, 322–352.

Shafer, G. (1992). Rejoinder to comments on 'Perspectives in the theory and practice of belief functions'. *International Journal of Approximate Reasoning*, *6*, 445–480.

Shafer, G. (1996). *The art of causal conjecture*. MIT Press.

Shenoy, P. (1989). A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, *3*, 383–411.

Shenoy, P. (1993). Valuation networks and conditional independence. In *Uncertainty in Artificial Intelligence (UAI'93)* (p. 191-199).

Shenoy, P. (1994). Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, *10*, 203–234.

Shenoy, P. (2000). Valuation network representation and solution of asymmetric decision problems. *European Journal of Operational Research*, *121*(3), 579-608.

Shenoy, P., & Shafer, G. (1990). Axioms for probability and belief functions propagation. In *Uncertainty in Artificial Intelligence (UAI'90)* (pp. 159–198).

Simon, C., Weber, P., & Evsukoff, A. (2008). Bayesian networks inference algorithm to implement Dempster-Shafer theory in reliability analysis. *Reliability Engineering and System Safety*, *93*, 950-963.

Smets, P. (n.d.). About updating. In *Uncertainty in Artificial Intelligence (UAI'91)* (pp. 378–385).

Smets, P. (1978). *Un modèle mathématico-statistique simulant le processus du diagnostic médical*. Ph.d. thesis, Université Libre de Bruxelles, (available through University Microfilm International, London).

Smets, P. (1988a). Belief functions. In *Non Standard Logics for Automated Reasoning* (pp. 253–286).

Smets, P. (1988b). The transferable belief model for quantified belief representation. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (Vol. 1, pp. 267–301).

Smets, P. (1990a). The combination of evidence in the transferable belief model. *IEEE Pattern Analysis and Machine Intelligence*, *12*, 447–458.

Smets, P. (1990b). Constructing the pignistic probability function in a context of uncertainty. In *Uncertainty in Artificial Intelligence (UAI'90)* (pp. 29–40).

Smets, P. (1992). Resolving misunderstandings about belief functions: A response to the many criticisms raised by judea pearl. *International Journal of Approximate Reasoning*, *6*, 321–344.

Smets, P. (1993a). An axiomatic justifiaction for the use of belief function to quantify beliefs. In *International Joint Conference on Artificial Intelligence (IJCAI'93)* (pp. 598–603).

Smets, P. (1993b). Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, *9*(1), 1-35.

Smets, P. (1993c). Jeffrey's rule of conditioning generalized to belief functions. In *Uncertainty in Artificial Intelligence (UAI'93)* (pp. 500–505).

Smets, P. (1995a). The canonical decomposition of a weighted belief. In *International Joint Conference on Artificial Intelligence (IJCAI'95)* (p. 1896-1901).

Smets, P. (1995b). Probability, possibility, belief: which for what? In *Foundation and Applications of Possibility Theory (FAPT'95)* (pp. 20–40).

Smets, P. (2000). Quantified possibility theory seen as an hyper cautious transferable belief model. In *Rencontres Francophones sur la Logique Floue et ses Applications (LFA'00)* (pp. 343–353).

Smets, P., & Hsia, Y. (1991). Default reasoning and the transferable belief model. In *Uncertainty in Artifical Intelligence (UAI'91)* (p. 495-505).

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (Vol. 1). MIT Press.

Spohn, W. (1988). Ordinal conditional functions : A dynamic theory of epistemic states. In *Causation in decision, belief changes and statistics* (pp. 105–134).

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453-489.

Suppes, P. (1970). *A probabilistic theory of causality.*

Teng, C. M. (2012). Applications of causal inference. In *International Symposium on Artificial Intelligence and Mathematics (ISAIM'12).*

Tian, J., & Pearl, J. (2002). *On the identification of causal effects* (Tech. Rep. No. R-290-L). UCLA C.S. Lab.

Trabelsi, S., Elouedi, Z., & Lingras, P. (2011). Classification systems based on rough sets under the belief function framework. *International Journal of Approximate Reasoning*, *52*(9), 1409-1432.

Tupin, F., Bloch, I., & Maître, H. (1999). A first step toward automatic interpretation of sar images using evidential fusion of several structure detectors. *IEEE T. Geoscience and Remote Sensing*, *37*(3), 1327-1343.

Vejnarová, J. (2012). Conditioning in evidence theory from the perspective of multidimensional models. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'12)* (Vol. 299 CCIS, p. 450-459). Springer-Verlag.

Verma, T., & Pearl, P. (1990). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence (UAI'90)* (p. 255-270).

Vogt, W. (2005). *Dictionary of statistics and methodology : a nontechnical guide for the social sciences.* Sage Publications.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford University Press.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557-585.

Xu, H., & Smets, P. (1994). Evidential reasoning with conditional belief functions. In *Uncertainty in Artificial Intelligence (UAI'94)* (p. 598-606).

Xu, H., & Smets, P. (1996). Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, *14*, 155–185.

Zadeh, L. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*, 3–28.

Zadeh, L. (2001). *Causality is undefinable* (Tech. Rep.). Univ. of California, Berkley.