Interventions and belief change in possibilistic graphical models

Salem Benferhat a,b,c,*

a Université Lille-Nord de France, F-59000, France
b Université d’Artois, CRIL, F-62307 Lens, France
c CNRS, UMR 8188, F-62307 Lens, Rue Jean Souvraz SP 18, 62307 Lens Cedex, France

Abstract

Causality and belief change play an important role in many applications. This paper focuses on the main issues of causality and interventions in possibilistic graphical models. We show that interventions, which are very useful for representing causal relations between events, can be naturally viewed as a belief change process. In particular, interventions can be handled using a possibilistic counterpart of Jeffrey’s rule of conditioning under uncertain inputs. This paper also addresses new issues that are arisen in the revision of graphical models when handling interventions. We first argue that the order in which observations and interventions are introduced is very important. Then we show that in order to correctly handle sequences of observations and interventions, one needs to change the structure of possibilistic networks. Lastly, an efficient procedure for revising possibilistic causal trees is provided.

1. Introduction

Causality and belief change are two important notions that play a crucial role in many applications in Artificial Intelligence and information systems. Both of them have been extensively studied in literature, in particular within different uncertainty theory frameworks (e.g., probability theory or possibility theory). The standard probabilistic definition of $A$ causes $B$, originally discussed in [1,2], requires that the probability degree of $B$ increases in the light of the new information $A$ (i.e., $Pr(B|A) > Pr(B)$). Namely, probabilistic causality is viewed as simple probability changes. This definition is symmetric and indirected which is not a desirable property in causality ascription. In order to overcome this limitation and make in evidence that causes precede their effects, other definitions of causality have been proposed (e.g., [3]). These definitions implicitly integrate the notion of “time” (of the occurrence of the events) in the characterisation of causality relation. Other definitions that take into account the fact that correlated events may be effects of some common cause (called a spurious cause in [3]) have also been proposed.

All these earlier definitions of probabilistic causation are given when uncertain information are explicitly described by means of probability distributions. Recently, Pearl [4] has proposed approaches based on probability theory to represent causal relations in graphical models. The proposal made by Pearl [4], called causal Bayesian networks, goes beyond standard Bayesian networks, where directed arcs in the graph are interpreted as representing elementary causal relations between variables in some real domain. Causal Bayesian networks also provide formal semantics to the notion of interventions, which plays an important role in causal ascription and for eliciting causal relations between variables. Interventions are external actions [5,4,6] that force some variables to have some specific values. From representational point of view, interventions...
are distinguished from observations using the concept of the "do" operator [7]. From reasoning point of view, handling interventions consists in "ignoring" the effects of all direct (and undirected) causes related to the variable of interest.

Causality notion in the possibilistic framework remains unexplored despite its importance. In fact, only a few works has addressed this issue (e.g., [8, 9]). This paper focuses on the representation and on the handling of interventions and causal information in possibilistic graphical models. In particular, we view the effects of interventions as a belief revision operation. The revision of a knowledge base or a database, consists of the insertion of some input information while preserving its consistency. During the past twenty years, many approaches have been proposed to address this problem from the axiomatic point of view (e.g., [10, 11]), from the semantics point of view (e.g., [12–14]) and from the computational point of view [15, 16]. In possibility theory, several revision operations have been proposed for both semantics and logical points of views. However, few works address the problem of a revision of possibilistic (or even probabilistic) networks.

This paper shows that interventions in possibilistic networks can be naturally described as a belief revision process. We provide natural properties that any possibilistic belief change operation should satisfy in order to correctly handle the presence of interventions. We then show that interventions have a very natural encoding in the possibilistic counterpart of Jeffrey's rule of conditioning under uncertain inputs. This paper also addresses new issues that are arisen in belief change when dealing with both observations and interventions. Indeed, we argue that the order in which observations and interventions are introduced matters: an observation followed by an intervention should not be equivalent to an intervention followed by an observation. Then we show that in order to respect this order then we need to change the structure of the graphical model after each observation and each intervention. An efficient procedure for revising possibilistic causal trees, in a presence of a sequence of observations and interventions, is then provided.

The rest of this paper is organised as follows. Section 2 provides a brief background on possibility theory. Section 3 recalls possibilistic networks and discusses its relationships with belief change. Section 4 is fully devoted to the handling of interventions in possibilistic graphical models. Section 5 concerns the problem of dealing with both observations and interventions, and propose an efficient procedure to revise possibilistic trees. Last section provides some concluding discussions.

2. Possibility theory: a refresher

2.1. Notations

Let \( V = \{A_1, A_2, \ldots, A_N\} \) be a set of variables. We denote by \( D_A = \{a_1, \ldots, a_n\} \) the domain associated with the variable \( A \). By \( a \) we denote any instance of \( A \). \( \Omega = \times_{A \in V} D_A \) denotes the universe of discourse, which is the Cartesian product of all variable domains in \( V \). Similarly, if \( X \subseteq \Omega \) is a set of variables, then \( D_X = \times_{A \in X} D_A \) denotes the set of instances of \( X \). Each element \( \omega \in \Omega \) is called an interpretation (or a state, an elementary event, etc.) of \( \Omega \). \( \omega[x] \) denotes the instances of the variables set \( X \) in \( \omega \). Similarly, \( [x] \) denotes the set of all interpretations where the variables \( X \) have their value equal to \( x \). Each interpretation of \( [x] \) is called a model of \( x \). An interpretation \( \omega \) in \( [x] \) is said to be a preferred model of \( x \) if there is no interpretation \( \omega' \) in \( [x] \) such that \( \pi(\omega') > \pi(\omega) \). If \( x \) and \( y \) are respectively instances of \( X \subseteq \Omega \) and \( Y \subseteq \Omega \), \([x \land y] \) denotes the set of all interpretations where \( X = x \) and \( Y = y \). In the following, in order to avoid heavy notations, and when there is no ambiguity we simply write \( x \) instead \( [x] \).

Next subsection only gives a brief recalling on possibility theory, for more details see [17–19].

2.2. Possibility distributions and possibility measures

A possibility distribution \( \pi \) is a mapping from the set of interpretations \( \Omega \) to the unit interval \([0, 1]\). It represents a state of knowledge about a set of possible interpretations distinguishing what is plausible from what is less plausible.

The value \( \pi(\omega) \) expresses a degree of compatibility (or coherence) of the interpretation \( \omega \) with respect to available knowledge encoded by \( \pi \). By convention, \( \pi(\omega) = 0 \) means that the interpretation \( \omega \) is impossible, and \( \pi(\omega) = 1 \) means that nothing prevents \( \omega \) from being the real world. When \( \pi(\omega) > \pi(\omega') \), \( \omega \) is a preferred candidate to \( \omega' \) for being the real state of the world. \( \pi \) is thus a convenient encoding of a preference relation that can embody concepts such as normality, typicality, plausibility, consistency with available beliefs, etc.

A possibility distribution \( \pi \) is said to be normalized (or consistent) if \( \exists \omega \in \Omega \), such that \( \pi(\omega) = 1 \), namely there exists at least an interpretation which is completely coherent with the available beliefs.

Given a possibility distribution \( \pi \) defined on a universe of discourse \( \Omega \), we can define a mapping grading the possibility measure of an event \( \phi \subseteq \Omega \) by:

\[
\Pi(\phi) = \max_{\omega \in \phi} \pi(\omega).
\]

\( \Pi(\phi) \) expresses the degree of compatibility of the event \( \phi \) with available knowledge encoded by \( \pi \).

Another mapping, called a necessity measure and which is dual to possibility measure, can be defined as:

\[
N(\phi) = 1 - \Pi(\bar{\phi}),
\]

where \( \bar{\phi} \) is the complement of \( \phi \) with respect to \( \Omega \). It evaluates to what extent \( \phi \) is inferred from beliefs encoded by \( \pi \).
2.3. Possibilistic conditioning

In a possibilistic setting (as in probability theory), conditioning consists in modifying our initial knowledge, encoded by a possibility distribution $\pi$, by the arrival of a new sure piece of information $\phi \subseteq \Omega$. We assume that $\phi$ is not a contradiction (namely $\phi$ is not empty) and that $\Pi(\phi) > 0$.

There are two main definitions of possibilistic conditioning. The first definition, proposed in [20] and [17], is called min-based conditioning and is defined by:

$$\pi(\omega | \phi) = \begin{cases} 1 & \text{if } \omega \in \phi, \Pi(\phi) = \pi(\omega), \\ \pi(\omega) & \text{if } \omega \in \phi, \Pi(\phi) > \pi(\omega), \\ 0 & \text{if } \omega \notin \phi. \end{cases}$$ (1)

The second definition is called product-based conditioning and is defined by:

$$\pi(\omega | \phi) = \begin{cases} \frac{\pi(\omega)}{\Pi(\phi)} & \text{if } \omega \in \phi, \\ 0 & \text{otherwise}. \end{cases}$$ (2)

These two definitions are the two well-used definitions of conditioning in possibility theory (see [21] for more discussions). Both of them satisfy the equation of the form:

$$\pi(\omega) = \pi(\omega | \phi) \Diamond \Pi(\phi),$$

which is similar to Bayesian conditioning, $\Diamond$ is either min or product. Besides when $\Pi(\phi) = 0$ then $\pi(\omega | \phi) = 1$ for both definitions.

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. In this paper, for sake of clarity, we only use the product-based conditioning. Product-based conditioning is also called Dempster conditioning, since it is a specialized of Dempster’s rule of conditioning for evidence theory, see for more details [22,23,18].

3. Possibilistic graphs

Graphical models (e.g., probabilistic networks [24–26], possibilistic networks [27–30], valuation-based systems [31]) are compact representations of uncertainty distributions. Their success is due to their simplicity and their capacity of representing and handling independence relationships, which are important for an efficient management of uncertain pieces of information. They have been used in different applications, such as diagnosis problems or intrusion detections. This section first provides basic definitions of possibilistic networks, then briefly discusses its relationships with belief change.

3.1. Basic definitions

A product-based possibilistic network on a set of variables $V$, denoted by $N = (\pi_N, G_N)$, consists of:

- a graphical component, denoted by $G_N$, which is a DAG (Directed Acyclic Graph). Nodes represent variables and edges encode the influence relation between variables. The set of parents of a node $A$ is denoted by $U_A$ and $\mu_A$ denotes an instance of parents of $A$;
- a numerical component, denoted by $\pi_N$, which quantifies different links of the network. For every root node $A$ ($U_A = \emptyset$), uncertainty is represented by a priori possibility degree $\pi_N(a)$ for each instance $a \in D_A$, such that:

$$\max_a \pi_N(a) = 1. \quad (4)$$

For the rest of the nodes ($U_A \neq \emptyset$) uncertainty is represented by a conditional possibility degree $\pi_N(a | \mu_A)$ of each instance $a \in D_A$ and $\mu_a \in D_{U_A}$. These conditional possibility distributions should satisfy the following normalization condition:

$$\max_a \pi_N(a | \mu_A) = 1, \quad \text{for any } \mu_A. \quad (5)$$

In the following, possibility distributions $\pi_N$, defined on nodes level, are called local possibility distributions.

From the set of local conditional possibility distributions, one can define a unique global joint possibility distribution similar to the one proposed in probability theory (e.g., [32,33]).

**Definition 1.** Let $N = (\pi_N, G_N)$ be a possibilistic network. The joint or global possibility distribution associated with $N$ and denoted by $\pi_N$, is expressed by the following chain rule:

$$\pi_N(A_1, \ldots, A_N) = \prod_{i=1,\ldots,N} \pi(A_i | U_{A_i}). \quad (6)$$
Let us present an example of possibilistic network which will be used in the whole paper for illustrating main concepts of the paper.

**Example 1.** Let $\mathbb{N} = (\pi_\mathbb{N}, G_\mathbb{N})$ be a possibilistic network. The DAG $(G_\mathbb{N})$ associated with $\mathbb{N}$ is given in Fig. 1. The example concerns a description of knowledge regarding “causal” or “influence” relation between the used ingredients and a hotness of a meal. For sake of simplicity, we assume that the variable “Hotness of the meals” admits three values: “Very Hot” (or very spicy), “Hot” and “Not Hot”. The variable “Ingredients” also admits three values: Paprika, Espelette pepper (a chili pepper that is cultivated in the French commune of Espelette), and Harissa (a North African hot red sauce). Harissa is a typical spicy ingredient that is used in North of Africa, even if several restaurants in this region use Paprika or French Espelette pepper in the preparation of their meals (especially to satisfy European tourists).

For sake of simplicity, in this example, we assume that possibility degrees belong to the following uncertainty scale $\mathcal{L} = \{1, \alpha, \alpha^2, \alpha^3, \ldots, 0\}$, with $1 > \alpha > 0$.

Table 1 describes our priori knowledge on the used ingredients in North African meals. It corresponds to a priori possibility distribution on the node Ingredients and expresses the facts that Harissa is the most normal used ingredient, Espelette is less normal and Paprika is the least plausible ingredient. Similarly, Table 2 describes our priori knowledge on the hotness of meals on the basis of used ingredients. For instance, $\pi_\mathbb{N}($Hotness | Paprika$)$ encodes the fact that in the context where the used ingredient is Paprika, the instance Not Hot is the most normal one, while the two other instances Hot and Very Hot are less normal but not excluded.

### 3.2. Belief change in possibilistic networks

The revision of a database or a knowledge base, consists of the insertion of some input information while preserving its consistency. A belief revision in its simple form [10] is a process, denoted by $\ast$, that transforms a consistent set of propositional formulae denoted by $K$ (called a belief set), into a new consistent set of formulae, denoted by $K \ast \psi$. The
propositional formula $\psi$ represents the input having a status of a sure piece of information that should be integrated in $K \ast \psi$.

In possibility theory, the revision process can be viewed as a so-called “transmutation” [34] that modifies the ranking of interpretations, called a possibility distribution, so as to give priority to the input information. Different forms of possibilistic revision have been proposed to deal with different type of inputs (a simple propositional formula, an uncertain observation, a possibility distribution, etc.). Possibilistic revisions have been extensively studied from both semantics and logical points of views [16,35–37].

However, few works address the problem of the revision of possibilistic (or even probabilistic) networks. In fact, a simple form of belief revision exists by means of propagation algorithms in possibilistic networks. A possibilistic network represents a background knowledge and induces a unique possibility distribution generated by Eq. (3). In the presence of new input observations of the form $X = x$ (namely, the new value of $X$ are $x$), possibilistic networks provide efficient tools that compute $P_N(A|X = x)$ for any event $A \subseteq \Omega$. Namely, they provide efficient ways to compute the possibility degrees of any event taking into account new input observations. For instance, when all variables are boolean, propagation algorithms can determine the set of current and accepted beliefs (i.e., a belief set), defined by $K_N = \{\phi: \phi \subseteq \Omega, \Pi(\phi) > \Pi(\phi^\sim)\}$, induced by a possibilistic network $\Psi$. They can also compute the result of revising $K_N$ by a new observation $\psi$, defined by: $K_N \ast \psi = \{\phi: \phi \subseteq \Omega, \Pi(\phi|\psi) > \Pi(\phi^\sim|\psi)\}$. If the graph associated with a possibilistic network has particular structures (such as polytrees), then queries of the form whether an event $\phi$ belongs or not to $K_N \ast \psi$ can be efficiently answered in a polynomial time. Moreover, the order in which observations are introduced does not matter.

However, propagation algorithms do not proceed to changing the structure of possibilistic networks in order to take into account new input observations. In fact as it is pointed out in [38], in possibilistic and probabilistic networks frameworks, iterating belief change just means accumulating consistent observations (namely all observations should be somewhat plausible) and reasoning from them using the background knowledge. The input observation is not considered as a new piece of knowledge to be integrated in possibilistic networks, it more corresponds to information pertaining only to the case at hand. Hence, there is no need to change the background knowledge represented by a causal network. We will see later that if one wants to respect this order then we need to change the structure of the graphical model after each observation and intervention.

### 4. Interventions as a belief change process

#### 4.1. Intervention vs observation

Interventions are commonly viewed as external actions that force some variables to have some specific values. They play an important role in distinguishing causation from mere correlation. A (global) joint probability alone can help to determine correlated or independent events but cannot be used to determine causal relations. To palliate this limitation, interventions can be used to arbitrate between several causal structures that fit the correlation data equally well. This core notion was introduced in early works (e.g., [5]), but was given its most prominent role by Pearl [4], whose definition of “$A$ causes $B$” requires that the forced occurrence of $A$, by means of an intervention, increases the probability of the occurrence of $B$.

Interventions provide a natural way for understanding causation. Identifying a causal relationship between different elements of the system would be much easier if the agent can directly intervene in the manner of an experimenter and evaluate the effects of such manipulations.

Interventions and observations should not be confused as illustrated by the following example:

**Example 2.** Let us consider again Example 1, where we assume now that you order a typical North African meal, and let us consider two different situations.

- **Observation:** Assume that you get some observation (for instance by tasting the meal), where the meal is not hot. Clearly, this new information will change your beliefs on the used ingredients. Your new beliefs will be:

  \[
  \pi_{N1}(\text{Paprika}) > \pi_{N1}(\text{Espelette}) = \alpha > \pi_{N1}(\text{Harissa}) = \alpha^2.
  \]

  The new possibility distribution $\pi_{N1}$ is the result of revising $\pi_N$ by taking into account the new observation that the meal is not hot. In fact, $\pi_{N1}$, given in Table 3, is simply obtained by conditioning $\pi_N$ (given in Table 2) with the event $\text{Hotness} = \text{Not Hot}$ using Eq. (1).

- **Intervention:** Consider now another situation where we assume that before tasting the meal, a waiter adds to the meal “Naga Jolokia”, the Indian chili tested hottest in the world at 1,040,000 SHU (Scoville heat units).1 Clearly, the waiter action forces the variable “Hotness” to get the value “Very Hot”. Or similarly, assume that a waiter adds “kabylian berber oil” which is known to make meals soft, namely that forces the variable “Hotness” to get the value “Not Hot”.

  One thing is sure, in both cases, and without additional information, after waiter’s action there is no reason to change our initial beliefs on the used ingredients in the meal. Namely:

Table 3
Revised joint distribution conditioned by the event Hotness = Not Hot.

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>Hotness</th>
<th>$\pi_N$(Hotness, Ingredients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paprika</td>
<td>Very Hot</td>
<td>0</td>
</tr>
<tr>
<td>Paprika</td>
<td>Hot</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Paprika</td>
<td>Not Hot</td>
<td>1</td>
</tr>
<tr>
<td>Espelette</td>
<td>Very Hot</td>
<td>0</td>
</tr>
<tr>
<td>Espelette</td>
<td>Hot</td>
<td>1</td>
</tr>
<tr>
<td>Espelette</td>
<td>Not Hot</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Harissa</td>
<td>Very Hot</td>
<td>0</td>
</tr>
<tr>
<td>Harissa</td>
<td>Hot</td>
<td>$\alpha^2$</td>
</tr>
<tr>
<td>Harissa</td>
<td>Not Hot</td>
<td>$\alpha^2$</td>
</tr>
</tbody>
</table>

$\pi_{N2}(Harissa) = \pi_N(Harissa) = 1$

> $\pi_{N2}(Espelette) = \pi_N(Espelette)$

$ = \alpha > \pi_{N2}(Paprika) = \pi_N(Paprika) = \alpha^2.$

where $\pi_{N2}$ is the result of modifying initial possibility distribution $\pi_N$ (given in Table 2) after the intervention on the variable “Hotness”.

From the above example, interventions are clearly external actions (of the systems) that force variables to take some specific values. They should not be confused with observations. The distinction between observations and interventions is similar to the one used between belief revision [10] and updating [39]. In belief revision the new information concerns a static world, while in updating it concerns changes in the world brought about by some agent. Observations, encoded as possibilistic conditionings, have been largely considered in the past from semantics and logical points of views.

The following section discusses the handling of intervention in possibility theory framework.

4.2. Handling interventions

From the example of the previous section, interventions on a variable $A = a$ can be viewed as a belief change process that transforms a possibility distribution $\pi_N$, into a new possibility distribution $\pi_{N1}$. Of course, we assume that $A = a$ is somewhat plausible in $\pi_N$, namely $\Pi_N(a) > 0$.

Let $U_A$ be the set of direct causes (or parents of $A$ in $G_N$), and $D_{U_A}$ be the domain associated with parents of $A$. The revised possibility distribution should at least satisfy the two following conditions:

R1: the event $A = a$ is a sure piece of information, namely any event where $A$ is different from $a$ is considered as impossible, and

R2: beliefs on direct causes of $A$ are unchanged.

Clearly, these two constraints are not enough to uniquely determine $\pi_{N1}$, and hence additional properties, in order to capture the idea of minimal change, are needed. We propose the following four natural and minimal properties for defining $\pi_{N1}$ (the result of revising $\pi_N$ after an intervention that forces $A$ to be equal to $a$):

A1: $\forall u_i \in D_{U_A}, \Pi_{N1}(u_i \land a) = \Pi_N(u_i)$ and $\forall u_i \neq a, \Pi_{N1}(a_i) = 0$.
A2: $\forall u_i \in D_{U_A}, \forall \omega, \omega' \in [u_i \land a] \text{ if } \pi_N(\omega) \geq \pi_N(\omega') \text{ then } \pi_{N1}(\omega) \geq \pi_{N1}(\omega')$.
A3: $\forall u_i \in D_{U_A},$ if $\Pi_N(u_i) = \Pi_N(u_i \land a)$ then $\forall \omega \in [u_i \land a]: \pi_{N1}(\omega) = \pi_N(\omega)$.
A4: If $\pi_N(\omega) = 0$ then $\pi_{N1}(\omega) = 0$.

A1 is a simple formal rewritting of the two conditions R1 and R2 described above. A2 means that the new possibility distribution should preserve the previous relative order (in the wide sense) between models of each $u_i \land a$, where $u_i$ is an instance of parents of $A$. A stronger version of A2 can be defined:

A$'_2$: $\forall u_i \in D_{U_A}, \forall \omega, \omega' \in [u_i \land a] \text{ then: } \pi_{N1}(\omega) > \pi_{N1}(\omega') \text{ iff } \pi_N(\omega) > \pi_N(\omega')$.

A$'_2$ is very similar to the two well-known postulates CR1, CR2 proposed in [11] for iterated belief revision in case where we restrict to propositional language (namely when all variables admit two instances).

A3 means that when $a$ is already accepted in a given parent context $u_i$ (namely, $\Pi_N(u_i) = \Pi_N(u_i \land a)$), then no changes occur inside models of $u_i \land a$. In particular, if for all $u_i \in D_{U_A}$ one have $\Pi_N(u_i) = \Pi_N(u_i \land a)$ then $\pi_N$ should not be altered, namely $\pi_N = \pi_{N1}$.
\[ A_4 \] stipulates that impossible interpretations remain impossible after intervention. This postulate fully makes sense in belief revision (where input observations concern static worlds), however it may be questionable when dealing with interventions (unless \( \Pi_N(\omega) > 0 \)). For instance, assume that in our running example that we observe that the meal is very hot, which means \( \forall \omega \notin \{\text{Very Hot}\}, \Pi_N(\omega) = 0 \). Then, after this observation, a waiter adds “kabylian berber oil” which forces the variable Hotness = Not Hot. Clearly, in this case \( A_4 \) contradicts \( A_1 \text{--} A_3 \). This is why for any intervention \( A = a \), we assume that \( \Pi_N(\omega) > 0 \).

Note that there are no further constraints which relate models of different \([u_i \land a]'s\) in the new possibility distribution \( \Pi_{N1} \). Note also that \( A_1 \text{--} A_4 \) imply that \( \Pi_{N1} \) is normalized. This is directly obtained from \( A_1 \) and from the fact that \( \max_{u_i \in DU_a} \Pi_N(u_i) = 1 \).

Clearly, properties \( A_1 \text{--} A_4 \) do not guarantee a unique definition of \( \Pi_{N1} \). In fact, in order to satisfy \( \forall u_i, \Pi_{N1}(u_i \land a) = \Pi_N(u_i) \), at least one of the preferred model of each \( u_i \land a \) should be upgraded until reaching \( \Pi_N(u_i) \). In order to avoid any arbitrary choice, a reasonable assumption is to consider that the possibility degrees of all best models of \( u_i \land a \) in \( \Pi_N \) should be increased to \( \Pi_N(u_i) \). Regarding remaining models \( u_i \land a \), we have two options:

- The first option is to simply left the possibility degrees of each non-preferred model unchanged. The possibility distribution \( \Pi_{N1} \) is then:
  \[
  \forall u_i, \forall \omega \in [u_i \land a], \quad \Pi_{N1}(\omega) = \begin{cases} 
  \Pi_N(u_i) & \text{if } \Pi_N(u_i \land a) = \Pi_N(\omega), \\
  \Pi_N(\omega) & \text{otherwise}
\end{cases}
  \tag{4}
\]
  and \( \forall \omega \notin [a], \Pi_{N1}(\omega) = 0 \).

- The second option is to proportionally rescale all models of \([u_i \land a]'s\) upwards, which leads to the following definition:
  \[
  \forall u_i, \forall \omega \in [u_i \land a], \quad \Pi_{N1}(\omega) = \frac{\Pi_N(u_i) \times \Pi_N(\omega)}{\Pi_N(u_i \land a)},
  \tag{5}
\]
  and \( \forall \omega \notin [a], \Pi_{N1}(\omega) = 0 \).

One can check that Eq. (5) (resp. Eq. (4)) can be obtained by using the possibilistic counterparts of Jeffrey’s rule [40] of conditioning under a particular set of uncertain inputs. The possibilistic counterpart of Jeffrey’s rule [40] provides an effective means to revise a possibility distribution \( \pi \) to \( \pi' \) given an input of the form of a possibility distributions bearing on a set of mutually exclusive and exhaustive events \( \phi_i 's \). More precisely, the input information is given in the form of pairs \( I = \{(\phi_i, \beta_i): i = 1, \ldots, n\} \), where \( \phi_i 's \) define a partition of \( \Omega \), with the two constraints:

\[
\forall (\phi_i, \beta_i) \in I, \quad \Pi'(\phi_i) = \beta_i
  \tag{6}
\]

and

\[
\forall (\phi_i, \beta_i) \in I, \quad \forall \psi \subseteq \Omega, \quad \Pi(\psi | \phi_i) = \Pi'(\psi | \phi_i).
  \tag{7}
\]

The underlying interpretation of revision implied by the constraint of Eq. (7) is that the revised possibility distribution \( \pi' \) must not change conditional probability degrees of any event \( \psi \) given any uncertain events \( \phi_i \).

When one uses product-based conditioning in Eq. (7), the two constraints above uniquely determine the possibility distribution \( \pi' \) given by:

\[
\forall (\phi_i, \beta_i) \in I, \forall \omega \in \phi_i, \quad \pi'(\omega) = \beta_i \times \pi(\omega | \phi_i).
  \tag{8}
\]

Clearly, Eq. (5) is a particular case of Eq. (8). It is enough to define the input \( I \) as:

\[
I = \{ ([u_i \land a], \Pi_N([u_i])]: u_i \in DU_A \text{ and } a \in DA \} \cup \{(a_i, 0): a_i \in DA \text{ and } a_i \neq a \}.
\]

Therefore, interventions can be naturally encoded using possibilistic counterpart of Jeffrey’s rule. This result reinforces the expressive power of possibilistic counterparts of Jeffrey’s rule of conditioning, since it has already been shown in [41] that it covers main belief revisions operations such as: adjustment [12], natural belief revision [13,42], drastic belief revision [43–45], revision of an epistemic by another epistemic state [46], improvement operators [47], etc. This result also shows that interventions provide a nice example where the application of Jeffrey’s rule is fully meaningful.

### 4.3 Other representations of interventions

The effects of an intervention on a variable \( A \) implies that our beliefs on parents set of \( A \), given by means of probability or possibility distributions, will not be affected. Clearly, the handling of interventions is more motived by a graphical structure rather than by uncertainty frameworks. Therefore, this section shows that the different graphical representations of interventions proposed in probabilistic causal models have natural counterparts in possibilistic graphical models (see also [9]). In fact, the “do” operator proposed in [4] to encode interventions, was firstly introduced in [7] in the ordinal conditional
functions of Spohn [42] a model for uncertainty representation closely related to infinitesimal probabilities and possibility theory [21,48].

Let \( \mathbb{N} \) be a possibilistic network, and assume that we have an intervention that forces a variable \( A_i \) to take a given value \( a_{ij} \). Interventions can then be described graphically in causal networks in two equivalent ways. The first way consists in modifying \( \mathbb{N} \) by proceeding to the deletion of links from \( U_{A_i} \) pointing into \( A_i \). The obtained possibilistic network is called a mutilated possibilistic network, denoted by \( \mathbb{N}_{\text{mut}} \). The joint distribution associated with \( \mathbb{N}_{\text{mut}} \) is then obtained by using Eq. (3). Let \( \omega = (a_1, \ldots, a_n) \in \Omega \) where for each \( i = 1, \ldots, n, a_i \) denotes an instance of the variable \( A_i \). We have:

\[
\forall \omega \in \Omega, \quad \pi_{\text{mut}}(\omega) = \begin{cases} 
* & \text{if } \omega[A_i] = a_{ij} \text{ and } \omega[U_{A_j}] = \mu_{A_j}, \\
0 & \text{otherwise}.
\end{cases}
\]

As it is illustrated by the schema of Fig. 2 the joint distribution obtained using Eq. (9) is equivalent to the joint distribution obtained using Eq. (5) or Eq. (8). Namely, given a possibilistic network \( \mathbb{N} \), handling intervention can equivalently be achieved: i) either by computing the joint distribution associated with \( \mathbb{N} \) then applying the possibilistic counterpart of Jeffrey’s rule of conditioning, or ii) by computing the mutilated graph and its associated joint distribution.

The second equivalent way to graphically representing interventions consists in adding, for each variable that may undergo an intervention, a new variable denoted \( \text{DO}_{A_i} \). The variable \( \text{DO}_{A_i} \) is added as a new parent of \( A_i \), and controls the status of the variables of \( A_i \). It takes a value \( \text{do}_{A_i} \text{-noact} \) when no intervention is observed and a value \( \text{do}_{A_i} \) when an intervention occurs forcing \( A_i \) to take the value \( a_i \) (\( a_i \) belonging to the domain of \( A_i \)). The resulting graph is called augmented possibilistic networks. Interventions on a variable \( A \) in an initial possibilistic network \( \mathbb{N} \) is then equivalent to observation on the node \( \text{DO}_{A_i} \) in its associated augmented possibilistic network.

Note that the augmented graphs representation of interventions is very close to the so-called virtual evidence [26] proposed for dealing with uncertain observations in standard probabilistic networks. This is not very surprising since on one hand we have seen that interventions have natural encoding using Jeffrey’s rule of conditioning under uncertain inputs, and on the other hand it has been shown in [49] that there are one by one translations between Pearl’s virtual evidence method and Jeffrey’s rule of conditioning.

From a computational point of view, in possibility theory framework, the better option to compute the effect of interventions is to use augmented graphs, since it allows the reuse of existent propagation algorithms (such as possibilistic adaptations of junction trees) without any extra costs (see [9] for more details).

Existing propagation algorithms in possibilistic networks do not allow to take into account a sequence of observations and interventions. Indeed, they are handled as if one first have all interventions, followed by all observations. The following section argues that the order in which interventions and observations arrive matter, and propose a solution for possibilistic causal trees.

5. Handling both interventions and observations

It is well known that revising a possibility distribution \( \pi \), using possibilistic conditioning, by observations is a commutative operation. Namely, a revision by a first observation \( A = a \) followed by a second observation \( B = b \), is equivalent to the revision by \( B = b \) followed by \( A = a \), and is also equivalent to the revision by observing simultaneously \( A = a \) and \( B = b \).

Similarly, one can easily check that the order in which interventions arrive does not matter. Namely, the intervention on a variable \( A \) followed by an intervention on a variable \( B \), is equivalent to first have an intervention on a variable \( B \) followed by an intervention on \( A \).
Example 3. Let us continue again our simple example, where we consider two scenarios:

- In the first scenario (an observation followed by an intervention), we first taste the meal and observe that it is not hot. Then, after this observation we add “Kabylian berber oil” to the meal. Clearly, at the end we obtain that the most plausible state is that the used ingredient is Paprika (induced from first observations), and the meal is not hot.
- In the second scenario (an intervention followed by an observation), we first add “Kabylian berber oil” to the meal, then we observe that the meal is not hot (without surprise). In this second scenario, the most plausible state is that the used ingredient is Harissa (since it was the most plausible used ingredient before the intervention), and the meal is not hot.

The fact that the order in which observations and interventions matter is not very surprising, since a similar behaviour holds for Jeffrey’s rule where it is well known that it does not commute [50]. The question now is how to deal with a sequence of observations and interventions, without an explicit computation of revised joint distributions resulting from taking into account new observations and interventions. The idea is to compute possibilistic networks which represent counterparts of revised possibility distributions. The situation is immediate in a presence of an intervention, thanks to the mutilated graph presented in the previous section. The problem is more difficult in the presence of observation. The following subsection addresses this issue.

5.1. Conditioning = combination + normalization

In order to compute a revised possibilistic causal network that integrates a new observation $A = a$, we will view the process of conditioning as a sequence of two operations: i) a combination of initial possibility distribution with the one associated with the input $A = a$, and ii) in case of conflict, a normalization operation of the possibility distribution obtained after the combination step. This construction is only meaningful under the following very reasonable assumption: If $A$ causes $B$, and if, after introducing a given observation $C$, $A$ and $B$ are related, then $A$ still causes $B$.

Let $N_1$ be a possibilistic network, and $\pi_{N_1}$ be its associated possibility distribution obtained using Eq. (3). Define:

$$\forall \omega \in \Omega, \quad \pi_{A=a}(\omega) = \begin{cases} 1 & \text{if } \omega[A] = a, \\ 0 & \text{otherwise.} \end{cases}$$

(10)

Then, let us define the combination operation as follows:

$$\forall \omega \in \Omega, \quad \pi_{N_2}(\omega) = \pi_{N_1}(\omega) * \pi_{A=a}(\omega).$$

(11)

The possibility distribution $\pi_{N_2}$ is obtained from $\pi_{N_1}$ by considering any interpretation $\omega$, where the value of $A$ is different from $a$, as impossible, while leaving possibility degrees of models of $a$ unchanged.

Clearly, $\pi_{N_2}$ may be subnormalized, and its normalization operation is defined as follows:

$$\forall \omega \in \Omega, \quad \pi_{N_3}(\omega) = \frac{\pi_{N_2}(\omega)}{h(\pi_{N_2})},$$

(12)

where $h(\pi_{N_2}) = \max_{\omega \in \Omega} \pi_{N_2}(\omega)$.

Then one can easily check that product-based conditioning, given by Eq. (1), is indeed equivalent to a sequence of two operations: a combination operation followed by a normalization operation. Namely:

$$\forall \omega \in \Omega, \quad \pi_{N_1}(\omega|A = a) = \pi_{N_3}(\omega).$$

(13)

We now provide the graphical counterparts of combination and normalization operations. For sake of simplicity, we restrict ourself to possibilistic causal networks where DAG’s are trees.

5.2. Graphical counterpart of the combination operation

This subsection provides the graphical counterpart of the combination operation, illustrated by Fig. 2. Let us denote by $N_2$ the result of combining $N_1$ with the new observation $A = a$. $N_2$ is obtained from $N_1$ by proceeding to two modifications on the node $A$ and its parent (see Fig. 5). More precisely,
Definition 2. The possibilistic network $N_2$ is such that:

- its graph $G_{N_2}$ is obtained from $G_{N_2}$ by simply removing the arc from the parent of $A$ (here denoted $D$) to $A$,
- its local possibility distributions, defined on variables different from $A$ and $D$, are identical to the ones given on $N_1$.

Regarding the variables $A$ and $D$ the new local possibility distributions are defined as follows:

- $\forall a_i \in D_A$, $\pi_{N_2}(a_i) = \begin{cases} 1 & \text{if } a_i = a, \\ 0 & \text{otherwise}. \end{cases}$
- $\forall d_i \in D_A, \forall c_j \in D_C$ $\pi_{N_2}(d_i|c_j) = \pi_{N_2}(d_i|c_j) \ast \pi_{N_1}(a|d_i)$.

The new local possibility distribution on the variable $A$ confirms that only the instance $a$ is fully possible, while the other instances are impossible. The local distribution on the node $D$ (parent of $A$) is modified in order to guarantee that possibility degrees of models of $a$ are identical in $\pi_{N_1}$ and $\pi_{N_2}$. Besides, since the value of the variable of $A$ is now fully determined, there is no need to maintain the arc from the parent of $A$ (here $D$) to $A$.

One can easily check that:

$\forall \omega \in \Omega$, $\pi_{N_2}(\omega) = \pi_{N_1}(\omega) \ast \pi_{A=a}(\omega)$.

Hence, $N_2$ is indeed the graphical counterpart of the combination operation.

5.3. Graphical counterpart of the normalization process

After the combination step, if $\pi_{N_2}$ is normalized then $N_2$ exactly encodes the result of revising $N_1$ by the observation $A = a$. Now, it may happen that $\pi_{N_2}$ is subnormalized, which means that $a$ was not accepted in initial possibilistic causal network $N_1$. In this case, we need to renormalize $\pi_{N_2}$. Our goal consists then in computing a possibilistic network, denoted by $N_3$, such that:

$\forall \omega$, $\pi_{N_3}(\omega) = \frac{\pi_{N_2}(\omega)}{h(\pi_{N_2})}$. \hfill (14)

The network $N_3$ is constructed such that all of its local possibility distributions are normalized. $N_3$ is obtained by progressively normalizing local distributions for each variable. We first study the case where only the local possibility distribution on the root variable in $N_2$ is subnormalized. In this particular case, $N_3$ is defined as follows:

Definition 3. Let $N_2$ be the possibilistic network obtained from the combination step. Assume that only the root variable, denoted by $A$, is subnormalized. Let $\max_{a \in D_A}(\pi_{N_3}(a)) = \beta$ and $0 < \beta < 1$. Let us define $N_3$ such that:

- $G_{N_3} = G_{N_2}$,
- $\forall X, X \neq A, \pi_{N_3}(X|U_X) = \pi_{N_2}(X|U_X)$,
- $\forall a \in D_A, \pi_{N_3}(a) = \pi_{N_2}(a)/\beta$.

Then we can show, in this particular case, that:

$\forall \omega$, $\pi_{N_3}(\omega) = \frac{\pi_{N_2}(\omega)}{h(\pi_{N_2})}$.

Namely, $N_3$ encodes the result of normalizing $N_2$.

The question now is how to deal with a variable $A$ which is not a root. In this situation, we solve the problem by modifying local possibility distributions associated with the parent of the variable $A$. This modification does not change joint
Definition 4. Let $N_2$ be a possibilistic network obtained from the combination step. Let $A$ be a variable and $B$ its parent. Assume that there is an instance $b$ of $B$ such that: $\max_{b \in D_B} \pi_{N_2}(a | b) = \beta$ where $0 < \beta < 1$. Let us define $N_3$ where its graph is the same as the one of $N_2$ and its local possibility distributions associated with $N_3$ are defined as:

1. $\forall C \neq A, \forall C \neq B, \pi_{N_3}(C | U_C) = \pi_{N_2}(C | U_C)$.
2. $\forall a_j, \forall b_i$,

$$
\pi_{N_3}(a_j | b_i) = \begin{cases}
\frac{\pi_{N_2}(a_j | b_i)}{\beta} & \text{if } b_i = b, \\
\pi_{N_2}(a_j | b_i) & \text{otherwise}.
\end{cases}
$$

3. $\forall b_i, \forall \mu_{b_i}$,

$$
\pi_{N_3}(b_i | \mu_{b_i}) = \begin{cases}
\frac{\pi_{N_2}(b_i | \mu_{b_i}) \ast \beta}{\pi_{N_2}(b_i | \mu_{b_i})} & \text{if } b_i = b, \\
\pi_{N_2}(b_i | \mu_{b_i}) & \text{otherwise}.
\end{cases}
$$

The three conditions of the above definition allow to normalize local possibility distributions at a node $A$. The first condition says that possibility distributions associated with variables, different from $A$ and $B$ (the parent of $A$), remain unchanged. The second condition specifies that normalization only affects the variable $A$. Lastly, the third condition applies on the variable $B$ (parent of $A$) the reverse operation of normalization, which ensures the equivalence between joint distributions. Hence, Definition 4 allows to normalize local distributions of variables having one parent, without modifying the joint global distribution, namely:

$$
\forall \omega, \quad \pi_{N_3}(\omega) = \pi_{N_2}(\omega).
$$

Definitions 3 and 4 allow to easily determine the possibilistic networks associated to the normalization of $\pi_{N_2}$. It is enough to repetitively apply Definitions 3 and 4, from leaves to roots, until all local possibility distributions are normalized. The renormalization of subnormalized possibilistic network is achieved in a linear time with respect to the number of parameters (possibility degrees) in the graphs.

Example 4. Let us continue our running example by providing the possibilistic networks in the two scenarios given in Example 3. Let $N_1$ be the possibilistic network given in Example 1, representing initial knowledge and beliefs before any interventions or observations.

- Recall that in the first scenario, we first have an observation that the meal is not hot, followed by the intervention of adding kablyan berber oil to the meal. In order to compute the possibilistic network resulting for the observation, we first apply Definition 2, which gives a possibilistic network $N_1$ composed of two independent nodes, and where its associated possibility distributions are described in Table 4.

The possibility distribution on the node “ingredient” is not normalized. Applying Definition 3 gives a possibilistic network $N_2$ where all of its associated possibility distributions are now normalized and are described by Table 5. One can easily check that the possibility distribution on the variable ingredients given in Table 5 is exactly the same as the one obtained from the result of conditioning the joint distribution (given in Table 2) with the event Hotness = Not Hot.
Lastly, the intervention has as no additional effect since the variable “Hotness” is now a root node and that Hotness = Not Hot is already accepted.

Clearly, Table 5 recovers the expected result, namely after the observation Hotness = Not Hot, followed by an intervention Hotness = Not Hot, we get that in the most plausible interpretation, the used ingredient is “Paprika” and that the meal is not hot.

- In the second scenario, we recall that we first have an intervention Hotness = Not Hot, followed by an observation Hotness = Not Hot. The handling of this second scenario is very simple since i) starting with an intervention immediately provides a possibilistic network N(2) with two independent nodes, with the possibility distributions \( \pi(N_3) \) given in Table 6, and ii) the observation has no effect after the intervention.

Again, in this scenario, we get the expected result where the most plausible event is that the used ingredient is “Harissa” and that the meal is not hot. This running example clearly illustrates that the order in which interventions and observations are introduced matter.

### 6. Concluding discussions

This paper dealt with the problem of handling interventions in possibilistic causal networks which play an important in representing causality relations and for causal ascription. We showed that interventions in possibilistic graphical models have a natural encoding as a belief change process, using the possibilistic counterpart of Jeffrey's rule. We provided several natural properties that a possibilistic belief change operation should satisfy in order to deal with interventions. We showed that in order to correctly deal with sequences of observations and interventions one needs to revise the structure of possibilistic networks. This clearly raises new issues in the problem of revising possibilistic networks, in particular how to revise a possibilistic causal networks when an intervention concerns an impossible event. In [8], the authors proposed a possibilistic model for causality ascription that answers queries of the form “is the event A a cause of the event B?” The proposed model is based on a non-monotonic inference relation which satisfies the requirements of System P [51]. Results of this paper can improve this model regarding at least three aspects. The first one is that using a possibilistic graphical model is based on a non-monotonic inference relation which satisfies the requirements of System P [51]. Results of this paper can improve this model regarding at least three aspects. The first one is that using a possibilistic graphical model allows to derive more causal relations, due to the addition of rational monotony property satisfied by a graphical model. The second aspect concerns the introduction of intervention in their model. Lastly, the third aspect concerns computational issues, where possibilistic graphical models offer efficient tools for reasoning from causal relations.

### Acknowledgements

This work is supported by the project ANR Blanc MICRAC.

### References


S. Benferhat / Artificial Intelligence 174 (2010) 177–189