Modeling positive and negative information in possibility theory *

Salem Benferhat, Didier Dubois, Souhila Kaci and Henri Prade
Institut de Recherche en Informatique de Toulouse (I.R.I.T.)–C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne 31062 Toulouse Cedex 4, France
E-mail:{benferhat, dubois, kaci, prade}@irit.fr

July 30, 2002

Abstract

From a representation point of view, it may be interesting to distinguish between i) what is possible because it is consistent with the available knowledge on the one hand, and ii) what is guaranteed to be possible because it is reported from observations on the other hand. Such a distinction also makes sense when expressing preferences, in order to point out choices which are positively desired among those which are not rejected. Possibility theory provides a representation framework where this distinction can be made in a graded way. In logical terms, the two types of information can be encoded by two types of constraints expressed in terms of necessity measures and in terms of guaranteed possibility functions. These two set functions are min-decomposable with respect to conjunction and disjunction respectively. This gives birth to two forms of possibilistic logic bases, which can viewed as generalized CNF and DNF respectively. By application of a minimal commitment principle, they induce a pair of possibility distributions at the semantic level, for which a consistency condition should hold in order to ensure that what is claimed to be guaranteed as possible is indeed not impossible. The bilattice structure underlying the framework is pointed out. The paper provides a survey of this bipolar representation framework, including the use of conditional measures, or the handling of comparative context-dependent constraints. The interest of the framework is stressed for expressing preferences, or in the representation of if then rules in terms of examples and counter-examples.

*This present draft borrows a large part of its contents (Sections 1 to 6) to the UAI-02 conference paper "Bipolar possibilistic representations" by the same authors.

1 Introduction

The idea of equipping a set of possible worlds, such as the interpretations of a classical logic language, with a complete pre-order is simple and useful for encoding preferences or plausibility orderings. This idea has emerged in the late eighties for providing non-monotonic consequence relations of the type "from p, q normally follows" requiring that all the most preferred/plausible/normal models of p be models of q [27]. Interestingly enough, possibility theory [30, 12] has been developed in the meantime, as a framework for modelling states of partial ignorance on the basis of possibility distributions encoding complete pre-orders. Then, possibility measures of events, defined as the maximum of the distribution over the set of models of the considered events, together with the dual measures of necessity, are used for assessing uncertainty. As such, possibility theory relies on a minimal specificity principle which states that the possibility level associated to a possible world should be the largest one which is compatible with the constraints representing the available information. It agrees in particular with the classical view of logical modelling according to which each of piece of information declares some worlds impossible (the others forming a subset of tentatively possible ones). When pieces of information are conjunctively combined, the resulting set of the possible worlds is indeed the largest set compatible with all pieces of information. Possibility theory extends this view by allowing intermediary levels between what is fully possible and what is completely impossible.

More recently, the idea that there exist situations where the information could be bipolar has been advocated in [29], and in [10] for the possibility theory framework. It corresponds to situations where two complete preorders, rather than one, are necessary to represent the whole information conveniently. While a consistency condition should be maintained between them, they behave in opposite ways in order to be complementary. Namely, while one of the two distributions agrees with the classical logical view mentioned above, the other rather corresponds to an idea of close world assumption, where only the information explicitly reported is held for true. The latter view obeys to a maximal specificity principle where only the worlds which are known as existing are considered to be possible. The information is then accumulated disjunctively in the sense that the larger the number of reported cases, the larger the range of worlds which are guaranteed to be possible.

This bipolar framework can be of interest for modelling knowledge in order to distinguish between what is possible because it is just consistent with what is known, and what is possible for sure since it has been observed. For instance, the information about the prices of houses in some area is usually partly made of general rules constraining the possible prices and of examples of reported cases. The bipolar view is also natural when modelling preferences if we have to distinguish between positive desires and preferences reflecting what it is not rejected as unsatisfactory [4]. Another situation which can lead to a bipolar

representation is the case where we start with a possibility distribution over a power set of possible worlds that we want to approximate from above and from below by distributions on the referential set.
The paper aims at showing how the possibility theory framework is convenient for handling bipolar representations. The next section explains how a pair of consistent distributions, referred to as lower and upper distributions respectively, should be associated with a treatment respectively in terms of guaranteed possibility functions (also called Δ-functions) and maximal specificity principle, and in terms of (classical) possibility measure Π and minimal specificity principle. A joint treatment of Δ and Π functions is provided, applying maximal specificity principle on Δ, and minimal specificity principle on Π. It is shown that the two principles are the two sides of a unique principle of minimal commitment. Section 3 summarizes logical representation results in terms of Δ-functions-based constraints and of necessity-based constraints. Sections 4, 5 and 6 show how the idea of conditioning, in terms of context-dependent comparative constraints, or in terms of conditional functions, can be applied to the bipolar representation framework. Sections 7 and 8 provide some illustrations of the usefulness of the proposed representation frameworks in reasoning with (fuzzy) rules, or for modelling preferences.

2 Bipolar possibility theory

2.1 Constraints on a possibility distribution

The basic representational tool in possibility theory is the notion of a possibility distribution [30], denoted by π, which encodes a complete preorder on a referential set U of possible worlds or possible interpretations. A degree $\pi(u) \in [0, 1]$ evaluates to what extent $u \in U$ can be the actual state of the world. So the larger $\pi(u)$, the greater the plausibility of u, the more normal u is, the more preferred u is, the more satisfactory u is, according to the problem under concern (knowledge or preference modelling).

A possibility distribution, representing the available information, should obey a *principle of minimal specificity*. This principle states that each possible interpretation u should receive the largest possibility degree in agreement with the set of constraints induced by the available information. If the available information states that "the actual state of the world is in a (classical) subset A", this means that any possibility distribution π such that $\forall u, \pi(u) \leq A(u)$ (we use the same notation for a subset and its characteristic function) is in accordance with the information. The information $\pi \leq A$ only rules out the values outside A which are asserted as completely impossible (since $A(u) = 0$ iff $u \notin A$). Choosing a particular π such that $\pi < A$, for representing the information, would be arbitrarily too precise. Hence choosing the largest possibility distribution π^* such that $\pi \leq A$, i.e. $\pi^* = A$, is natural. Clearly, a collection of constraints $\pi \leq A_i$ for $i = 1, \cdots, n$ should be aggregated conjunctively, i.e.,$\pi \leq min_i A_i$. The more constraints of this form, the more precise our information about the location of the actual state of the world. It corresponds to the situation in classical logic, where the more formulas, the smaller the set of interpretations which make them true. The full consistency of the constraints is achieved when π^* is *normalized*, i.e. $\exists u, \pi^*(u) = 1$, which means that there remains at least one fully possible interpretation.

This approach extends to the case where the A_i's or A are fuzzy subsets. Let $A_{\overline{\alpha}} = \{u : A(u) > \alpha\}$. Note that the $A_{\overline{\alpha}}$'s are nested, i.e. $\alpha \geq \beta \Rightarrow A_{\overline{\alpha}} \subseteq A_{\overline{\beta}}$. Then $\pi \leq A$ means that $\forall\alpha, \forall u \notin A_{\overline{\alpha}}, \pi(u) \leq \alpha$, i.e. the possibility of u is upper-bounded by α. This can be translated in terms of the set of constraints $\forall\alpha, \Pi((A_{\overline{\alpha}})^c) \leq \alpha$, where A^c is the complement of A in U and where Π is a *possibility measure* defined from π by

$$\Pi(A) = sup_{u\in A}\pi(u), \forall A \subseteq U.$$

It can be as well stated in terms of the dual necessity measure $N(A) = 1-\Pi(A^c)$, i.e.

$$\forall\alpha, N(A_{\overline{\alpha}}) \geq 1 - \alpha, \quad (1)$$

which expresses that we are certain, at least at level $1-\alpha$, that the actual world is in $A_{\overline{\alpha}}$.

But there exists a converse type of constraint of the form $\pi \geq A$, which expresses that all the values in A are possible at least at the degree to which these values belong to A. It corresponds to a different kind of information of the type "all the values in A are possible". Such a piece of information is natural when reporting on observed data whose feasibility, or possibility of appearance is thus guaranteed. The aggregation of such pieces of information is disjunctive since $\pi \geq A_j$ for $j = 1, \cdots, m$ entails $\pi \geq max_j A_j$ and corresponds to a data accumulation process. The notion of possibility underlying the constraint $\pi \geq A$ is related to a set function, denoted Δ, different from the possibility measure Π. Indeed $\pi \geq A$ means that $\forall\alpha, \forall u \in A_\alpha, \pi(u) \geq \alpha$, i.e.

$$\forall\alpha, \Delta(A_\alpha) \geq \alpha, \quad (2)$$

where $A_\alpha = \{u : A(u) \geq \alpha\}$ and $\Delta(A) = inf_{u\in A}\pi(u)$ is the minimum degree of possibility over A and is called the *"guaranteed possibility"* of A.

Clearly with constraints of the form $\pi \geq A$, we cannot apply a principle of minimal specificity (which would lead to the state of complete ignorance, $\forall u, \pi(u) = 1$). We should rather use the converse *principle of maximal specificity* which allocates to each value its minimum degree of possibility enforced by the constraint. It corresponds to the idea of focusing only on what has been actually observed. So the information $\pi \geq A$ translates into $\pi_* = A$ applying this principle. In particular $\pi_*(u) = 0$ if $u \notin \{u : A(u) > 0\}$, since then there is no reason to assume that u is (somewhat) possible, as far as it has not been reported.

Note that Π is a maxitive set function (increasing in the wide sense with set inclusion), i.e. $\Pi(A \cup B) = max(\Pi(A), \Pi(B))$, $\forall A, \forall B$, while guaranteed possibility functions are decreasing set functions such that $\Delta(A \cup B) = min(\Delta(A), \Delta(B))$. This means that interpretations covered by A or B are guaranteed to be possible (i.e., because they are observed, feasible, satisfactory, permitted, according to the problems) if and only if both the interpretations in A and those in B are guaranteed to be possible.

2.2 The bipolar view

Generally, the two types of information described above may be available, namely the ones of the form $\pi \leq A_1$ and those of the form $\pi \geq A_2$, leading to the obvious *consistency* requirement $\pi_* = A_2 \leq \pi^* = A_1$, i.e.,

$$\forall u, \pi_*(u) \leq \pi^*(u). \tag{3}$$

This agrees with the idea that what is observed is only a part of what is possible.

Thus a pair of consistent lower and upper possibility distributions π_* and π^*, obeying to (3) can be seen as encoding two types of information of the form respectively $\Delta(A_{*\alpha}) \geq \alpha$ where $\pi_* = A_*$, i.e. the worlds in $A_{*\alpha}$ are guaranteed to be possible at degree α on the one hand, and of the form $N(A^*_{\overline{\alpha}}) \geq 1 - \alpha$ where $\pi^* = A^*$, i.e. it is certain at degree $1 - \alpha$ that the actual world is in $A^*_{\overline{\alpha}}$. Given a consistent pair of possibility distributions π_* and π^* ($\pi_* \leq \pi^*$), the uncertainty of an event A is evaluated by the ordered pair $(\Delta(A), \Pi(A))$, with $\Delta(A) \leq \Pi(A)$, since Δ is defined from π_* and Π from π^*. This can be symbolically written $(\Delta, \Pi)(A) = (\Delta(A), \Pi(A))$. Then we have

$$(\Delta, \Pi)(A \cup B) = mM((\Delta, \Pi)(A), (\Delta, \Pi)(B)), \tag{4}$$

where m stands for min and M for max, and $mM((a, b), (c, d)) = (min(a, c), max(b, d))$. Introducing by duality

$$(N, \nabla)(A) = (N(A), \nabla(A)) = 1 - (\Delta, \Pi)(A^c) \tag{5}$$

$$= 1 - (\Delta(A^c), \Pi(A^c))$$
$$= (1 - \Pi(A^c), 1 - \Delta(A^c)),$$

where $1 - (a, b) = (1 - b, 1 - a)$, we have similarly

$$(N, \nabla)(A \cap B) = mM((N, \nabla)(A), (N, \nabla)(B)).$$

$\nabla(A)$ is called the *potential necessity* of A. Indeed, we have $N(A) \leq \nabla(A)$. Providing that π^* is normalized and that $\exists u, \pi_*(u) = 0$ (which is always possible by adding an extra element), the following inequality holds

$$max(N(A), \Delta(A)) \leq min(\Pi(A), \nabla(A)). \tag{6}$$

Up to (6), the four set functions N, Δ, Π and ∇ are unrelated. This strongly contrasts with the probabilistic setting where only one quantity assesses the uncertainty of A, since $Prob(A^c) = 1 - Prob(A)$.
In general, while (4) holds for the disjunction, we only have the inequality

$$\Pi(A \cap B) \leq min(\Pi(A), \Pi(B)),$$

for possibility measures, since $A \cap B$ may be impossible while $A \cap B^c$ and thus A can be fully possible, as well as $A^c \cap B$ and thus B. A counterpart of this inequality holds for Δ, namely

$$\Delta(A \cap B) \geq max(\Delta(A), \Delta(B)),$$

since the minimum of π_* over $A \cap B$ may be greater than the minima over A, and over B. These two inequalities can be written more compactly as

$$(\Delta, \Pi)(A \cap B) \subseteq Mm((\Delta, \Pi)(A), (\Delta, \Pi)(B)), \tag{7}$$

where $Mm((a, b), (c, d)) = (max(a, c), min(b, d))$ and where $(a, b) \subseteq (c, d)$ means $a \geq c$ and $b \leq d$. The partial order "$\subseteq$" could be understood here as meaning "better informed than". Note that the principles of minimal and maximal specificity jointly applied to upper and lower constraints $\pi^* \leq A_1$ and $\pi_* \geq A_2$, i.e. $(\pi_*, \pi^*) \subseteq (A_2, A_1)$, leading to take $(\pi_*, \pi^*) = (A_2, A_1)$, amounts to a *unique principle of minimal commitment* regarding the available information: do not represent more information than you have. The principle of minimal commitment applied to (7) leads to the equality

$$(\Delta, \Pi)(A \cap B) = Mm((\Delta, \Pi)(A), (\Delta, \Pi)(B)). \tag{8}$$

When (8) holds, $\Delta(A \cap B)$ (resp. $\Pi(A \cap B)$) does not bring further information w.r.t. $\Delta(A)$ and $\Delta(B)$ (resp. $\Pi(A)$ and $\Pi(B)$).
Observe that (8) encompasses the condition $\Pi(A \cap B) = min(\Pi(A), \Pi(B))$, which is a form of independence between A and B, called by Nahmias [26], "unrelatedness". The corresponding form of *independence* for Δ reads, $\Delta(A \cap B) = max(\Delta(A), \Delta(B))$, i.e. if A and B are unrelated, the interpretations in A and B are guaranteed to be possible as far as the interpretations of A, or those of B, are guaranteed to be possible.

Besides, apart from '$\subseteq$', there exists another natural partial ordering which can be introduced in the bipolar framework, that we shall denote by $\leq$, which means "*less possible than*", and which is defined by $(\Delta, \Pi)(A) \leq (\Delta, \Pi)(B)$ iff $\Delta(A) \leq \Delta(B)$ and $\Pi(A) \leq \Pi(B)$. These two orderings induce a *bilattice* structure in the sense of Ginsberg [18].

3 Bipolar possibilistic logic

Lower and upper possibility distributions are not always directly available, but are rather implicitly specified by means of guaranteed possibility-based con-

straints and necessity-based constraints respectively. These constraints bear on sets of possible worlds which are no longer necessarily nested, as in the case of the α-level cuts of the fuzzy sets specifying the distributions in Section 2.1.
In possibilistic logic [11], the information is prioritized in terms of necessity measures. A *possibilistic logic base* is a set of weighted formulas of the form $\Sigma = \{(a_i, \gamma_i) : i = 1, \cdots, n\}$, where a_i is a propositional formula (which corresponds to a set of models A_i) and γ_i belongs to a totally ordered scale such as $[0, 1]$. (a_i, γ_i) means that the necessity degree of A_i is at least equal to γ_i, i.e. $N(A_i) \geq \gamma_i$.
With a possibilistic logic base Σ, a unique possibility distribution, denoted by π_Σ, is associated at the semantical level; it is defined by [11]: $\forall u \in U$,

$$\pi^*_\Sigma(u) = \begin{cases} 1 & \text{if } \forall (a_i, \gamma_i) \in \Sigma, u \models a_i \\ 1 - max\{\gamma_i : (a_i, \gamma_i) \in \Sigma \text{ and } u \not\models a_i\} & \text{otherwise,} \end{cases}$$

where π^*_Σ is in agreement with the minimal specificity principle. Indeed this is the largest possibility distribution whose the associated necessity measure N satisfies $N(A_i) \geq \gamma_i$ for all i such that $(a_i, \gamma_i) \in \Sigma$.
Besides, the inference rule $(a \vee b, \alpha), (\neg a \vee c, \beta) \vdash (b \vee c, min(\alpha, \beta))$ is the basis of the possibilistic inference machinery at the syntactic level [11].
In possibilistic logic each formula (a_i, γ_i) expresses that it is certain at level γ_i that the actual state of the world is among the models of a_i, i.e. that the counter-models of a_i are only possible at most at level $1 - \gamma_i$. Then a possibilistic logic base $\Sigma = \{(a_i, \gamma_i) : i = 1, \cdots, n\}$ is associated with the upper possibility distribution π^*_Σ defined above.

Another type of information is encoded by constraints of the form $\Delta(B_j) \geq \delta_j$ for asserting that any model of a proposition b_j, whose set of models is B_j, is a possible candidate for being the actual world at least at level δ_j. Such a set of constraints defines what can be called a lower possibilistic base, of the form $\Xi = \{[b_j, \delta_j] : j = 1, \cdots, m\}$, where $[b_j, \delta_j]$ encodes the inequality $\Delta(B_j) \geq \delta_j$. Applying the maximal specificity principle, leads to derive the following lower distribution with Ξ, defined by

Definition 1

$$\pi_{*\Xi}(u) = \begin{cases} 0 & if \;\; \not\exists j, u \models b_j \;\; and \;\; [b_j, \delta_j] \in \Xi \\ max\{\delta_j : u \models b_j, [b_j, \delta_j] \in \Xi\} & otherwise. \end{cases}$$

$\pi_{*\Xi}$ is the smallest distribution whose associated Δ-function satisfies $\forall j, \Delta(B_j) \geq \delta_j$.
Besides, the logical machinery for Δ-weighted formulas is governed by the following cut rule $[a \wedge b, \alpha], [\neg a \wedge c, \beta] \vdash [b \wedge c, min(\alpha, \beta)]$ [10], which is the counterpart of the possibilistic resolution rule, exchanging Δ and N, $\wedge$ and $\vee$.

4 Bipolar conditioning

A conditional possibility measure $\Pi(. \mid A)$ has been defined by Hisdal [22] as satisfying the Bayes-like relationship:

$$\Pi(A \cap B) = min(\Pi(B \mid A), \Pi(A)). \tag{9}$$

The principle of minimal specificity leads to define $\Pi(. \mid A)$ as the greatest solution of this equation, namely, if $A \neq \emptyset$:

$$\Pi(B \mid A) = \begin{cases} 1 & \text{if } \Pi(A \cap B) = \Pi(A) \\ \Pi(A \cap B) & \text{if } \Pi(A \cap B^c) > \Pi(A \cap B). \end{cases}$$

$\Pi(B \mid A) = \Pi(B)$ implies $\Pi(A \cap B) = min(\Pi(A), \Pi(B))$, i.e. A and B are unrelated in the sense of Section 2 (not conversely). The dual conditional necessity measure is simply $N(B \mid A) = 1 - \Pi(B^c \mid A)$.
This is an ordinal form of conditioning, adapted to finite settings, with a qualitative (ordinal) possibility scale. The corresponding conditional possibility distribution $\pi(. \mid A)$ is defined by:

$$\pi(u \mid A) = \begin{cases} 1 & \text{if } u \in A \text{ and } \pi(u) = \Pi(A) \\ \pi(u) & u \in A \text{ and } \pi(u) < \Pi(A) \\ 0 & u \notin A. \end{cases}$$

Conditioning can be also defined for guaranteed possibility measures. As for the inference, it works in a reversed way w.r.t. Π. Namely, conditioning obeys the following equation

$$\Delta(A \cap B) = max(\Delta(B \mid A), \Delta(A)), \tag{10}$$

It means that $A \cap B$ is guaranteed possible either because A is so or because, in the context A, B is guaranteed possible.
Since $\Delta(A) = min(\Delta(A \cap B), \Delta(A \cap B^c))$, it expresses either that the minimum of π_* over A is reached on $A \cap B$ and that $\Delta(B \mid A)$ can be taken as equal to 0 by virtue of the principle of maximal specificity, or that $\Delta(A) = \Delta(A \cap B^c) < \Delta(A \cap B)$ and then $\Delta(B \mid A)$ should be equal to $\Delta(A \cap B)$. Thus, we have

$$\Delta(B \mid A) = \begin{cases} \Delta(A \cap B) & \text{if } \Delta(A) < \Delta(A \cap B) \\ 0 & \text{if } \Delta(A \cap B^c) \geq \Delta(A \cap B) = \Delta(A). \end{cases}$$

It can be checked that if $\Delta(B \mid A) = 0$ then A and B are unrelated, i.e. $\Delta(A \cap B) = max(\Delta(A), \Delta(B))$.
Interestingly enough, we can define upper and lower possibilistic conditioning in a compact way as obeying the relation

$$(\Delta, \Pi)(A \cap B) = Mm((\Delta, \Pi)(B \mid A), (\Delta, \Pi)(A)). \tag{11}$$

Moreover it can be shown that

$$N(B \mid A) > 0 \text{ iff } \Pi(A \cap B) > \Pi(A \cap B^c)$$

iff $N(A \Rightarrow B) > N(A \Rightarrow B^c)$, where $\Rightarrow$ denotes material implication. It expresses that B is somewhat certain in context A iff $A \cap B$ is strictly more plausible than $A \cap B^c$. A similar relation holds for guaranteed possibility: $\Delta(B \mid A) > 0$ iff $\Delta(A \cap B) > \Delta(A \cap B^c)$ iff $\nabla(A \Rightarrow B) > \nabla(A \Rightarrow B^c)$.
This points out the fact that the definition of conditioning enforces the following normalization condition:

$$min(\Delta(B \mid A), \Delta(B^c \mid A)) = 0.$$

This expresses that everything (i.e. B and B^c) cannot be guaranteed as being simultaneously possible, and the fact that B is guaranteed as being possible in context A means that we have more guarantees in favor of B than in favor of B^c.
Thus one can write $N(B \mid A) > 0$ and $\Delta(B \mid A) > 0$ iff $(\Delta, \Pi)(A \cap B) > (\Delta, \Pi)(A \cap B^c)$
using the second partial ordering introduced in Section 2.

5 Strict comparative possibility bases

A strict comparative possibility base $\mathcal{P}$ is a set of constraints of the form "in context a, b is more possible than $\neg b$" i.e. in terms of models, $\Pi(A \cap B) > \Pi(A \cap B^c)$, denoted by $a \to b$. It means that b is true in all best models of a. This can either express a general rule having exceptions, or the conditional preference of an agent. It encompasses the general case of constraints of the form $\Pi(A) > \Pi(B)$. Indeed $\Pi(A) > \Pi(B)$ is trivially equivalent to $\Pi(A \cap B^c) > \Pi(B)$ and corresponds to the rule $a \vee b \to \neg b$ since $A \cap B^c = (A \cup B) \cap B^c$.

In a qualitative setting, a possibility distribution π can be represented by its well-ordered partition $WOP(\pi) = E_1 \cup \cdots \cup E_n$ such that: $E_1 \cup \cdots \cup E_n = U$, $E_i \cap E_j = \emptyset$, and $\forall u, u', \pi(u) > \pi(u')$ iff $u \in E_i$, $u' \in E_j$ and $i < j$.

Each strict comparative possibility base $\mathcal{P}$ induces a unique upper qualitative possibility distribution $WOP(\pi_{\mathcal{P}})$ obtained by considering the least specific solution satisfying:

$$\Pi(A_k \cap B_k) > \Pi(A_k \cap B_k^c) \tag{12}$$

for all $a_k \to b_k$ of $\mathcal{P}$.
In [5], an algorithm has been provided to compute $WOP(\pi_{\mathcal{P}})$= $E_1 \cup \cdots \cup E_n$. Here we only recall its basic principle, which consists in putting each interpretation in the lowest possible rank (or highest possibility degree) without violating constraints (12). The only case where we cannot put u in some partition E_i, is when u is in the right part of some constraint (where there is a rule $a_k \to b_k$ such that $u \models a_k \wedge \neg b_k$), and none of the interpretations of the left part of this constraint (i.e., $u \models a_k \wedge b_k$) is already classified in some E_j with $j < i$. Therefore $WOP(\pi_{\mathcal{P}})$ is computed as follows: for each step i, we put in E_i all

interpretations which are not in the right part of any constraint, then we remove all rules $a_k \to b_k$ such that there exists at least u in E_i such that $u \models a_k \wedge b_k$.

There exists a converse transformation from π to $\mathcal{P}$, see [3].

Similarly, a set of Δ-based comparative constraints $\Delta(A_k \cap B_k) > \Delta(A_k \cap B_k^c)$ defines a lower qualitative possibility distribution by applying the maximal specificity principle.

Let $G = \{a_k \rightsquigarrow b_k : k = 1, \cdots, n\}$ be a set of rules. Let us denote by $\mathcal{C}= \{\Delta(A_k \cap B_k) > \Delta(A_k \cap B_k^c) : a_k \rightsquigarrow b_k \in G\}$ the set of Δ-based comparative constraints induced by G. Algorithm 1 provides the most specific qualitative distribution satisfying G, denoted by $WOP(\pi_G)$.

The idea of the algorithm consists in assigning to each interpretation the lowest

Algorithm 1:

Data: G
Result: $WOP(\pi) = E_1 \cup \cdots \cup E_n$
begin
    $i \leftarrow 0$;
    while *($U \neq \emptyset$)* **do**
        $i \leftarrow i + 1$;
        $S_i = \{u : \nexists \Delta(A_k \cap B_k) > \Delta(A_k \cap B_k^c)$ s.t. $u \in A_k \cap B_k\}$;
        if $S_i = \emptyset$ **then**
            G is inconsistent
        - Remove from U elements of S_i;
        - Remove from $\mathcal{C}$ constraints such that
        $S_i \cap (A_k \cap B_k^c) \neq \emptyset$
    return $E_1, \cdots, E_i$ *s.t.* $\forall j \leq i, E_j = S_{i-j+1}$
end

possibility degree. At each step i, we put in S_i the interpretations which are not in left part of any Δ-based constraint $\Delta(A_k \cap B_k) > \Delta(A_k \cap B_k^c)$ (otherwise such a constraint will be falsified). For instance, the least plausible interpretations are those which do not verify any Δ-rule (namely, are not in any $A_k \cap B_k$).

Example 1 *Let us consider an example where $G = \{p \rightsquigarrow q, p \rightsquigarrow \neg r, q \rightsquigarrow r\}$, and p, q, r are three propositional symbols. Let $U = \{\omega_0 = pqr, \omega_1 = pq\neg r, \omega_2 = p\neg qr$, $\omega_3 = p\neg q\neg r, \omega_4 = \neg pqr, \omega_5 = \neg pq\neg r$, $\omega_6 = \neg p\neg qr, \omega_7 = \neg p\neg q\neg r\}$.*
Let us apply our previous algorithm. We have:
$\mathcal{C}=$ $\Delta(\{\omega_0, \omega_1\}) > \Delta(\{\omega_2, \omega_3\})$, $\Delta(\{\omega_1, \omega_3\}) > \Delta(\{\omega_0, \omega_2\})$, $\Delta(\{\omega_0, \omega_4\}) > \Delta(\{\omega_1, \omega_5\})\}$.
The use of maximal specificity principle leads to first put $\omega_6, \omega_7, \omega_2$ and ω_5 in the lowest rank since they are not constrained, namely from our algorithm, we have: $S_1 = \{\omega_6, \omega_7, \omega_2, \omega_5\}$.
Once S_1 is settled, all the constraints in $\mathcal{C}$ are satisfied, hence S_2 contains the

remaining interpretations.
Hence, the partition associated to G is:
$E_1 = \{\omega_0, \omega_1, \omega_3, \omega_4\}$ *and* $E_2 = \{\omega_2, \omega_5, \omega_6, \omega_7\}$.

A syntactic counterpart of the above algorithm where G would be directly encoded in terms of Δ-based formulas, can be provided by taking advantages of transformations between the different representations of a possibility distribution [3]. Its computational complexity is not more costly than the one used to transform a strict comparative possibility base into a possibilistic logic base [5].

6 Possibilistic networks

Another compact representation of a possibility distribution is graphical and is based on conditioning. Symbolic knowledge is represented by DAGs, where nodes represent variables (in this paper, we assume that they are binary), and edges express influence links between variables. When there exists a link from X to Y, X is said to be a parent of Y. The set of parents of a given node X is denoted by $Par(X)$. By the capital letter X we denote a variable which represents either the symbol x or its negation. An interpretation in this section will be simply denoted by $X_1 \cdots X_n$. Uncertainty is expressed at each node by a pair of lower and upper possibility distributions as follows:
– For root nodes X_i we provide the prior possibility degrees and guaranteed possibility degrees $(\Delta(X_i), \Pi(X_i))$ of x_i and of its negation $\neg x_i$. These prior should satisfy the normalization condition: $max(\Pi(x_i), \Pi(\neg x_i)) = 1$. The normalization condition $min(\Delta(x_i), \Delta(\neg x_i)) = 0$ is assumed for guaranteed possibility degrees.
– For other nodes X_j, we provide $(\Delta(X_i \mid Par(X_i)), \Pi(X_i \mid Par(X_i)))$ i.e. the conditional possibility degrees and guaranteed possibility degrees of x_j and of its negation $\neg x_j$ given any complete instantiation of each variable of parents of X_j, $u_{Par(X_j)}$. The conditional possibilities should also satisfy the normalization conditions:

$$max(\Pi(x_j \mid u_{Par(X_j)}), \Pi(\neg x_j \mid u_{Par(X_j)})) = 1.$$
$$min(\Delta(x_j \mid u_{Par(X_j)}), \Delta(\neg x_j \mid u_{Par(X_j)})) = 0.$$

The local possibilities $(\Delta(X_i), \Pi(X_i))$ (resp. $(\Delta(X_i \mid Par(X_i)), \Pi(X_i \mid Par(X_i))$) should satisfy the coherence conditions:

$$\Delta(X_i) \leq \Pi(X_i)$$

(resp. $\Delta(X_i \mid Par(X_i)) \leq \Pi(X_i \mid Par(X_i))$).

A first approach to defining a pair of joint distributions associated with a bipolar possibilistic graph is to apply the following rules:

$$\pi^*(u) = min\{\Pi(x \mid u_{Par(X)}) : u \models x \text{ and } u \models u_{Par(X)}\}, \tag{13}$$

$$\pi_*(u) = max\{\Delta(x \mid u_{Par(X)}) : u \models x \text{ and } u \models u_{Par(X)}\}. \tag{14}$$

	x_1x_2	$x_1\neg x_2$	$\neg x_1x_2$	$\neg x_1\neg x_2$
x_3	$(.4, 1)$	$(0, 1)$	$(0, .5)$	$(.4, .8)$
$\neg x_3$	$(0, .6)$	$(.3, 1)$	$(.2, 1)$	$(0, 1)$

Table 1:

u	$\pi^*(u)$	$\pi_*(u)$
$\neg x_1\neg x_2\neg x_3$	.5	.45
$\neg x_1\neg x_2x_3$	.5	.45
$\neg x_1x_2\neg x_3$	1	.45
$\neg x_1x_2x_3$	.5	.45
$x_1\neg x_2\neg x_3$	.5	.4
$x_1\neg x_2x_3$	.5	.4
$x_1x_2\neg x_3$	.5	0
$x_1x_2x_3$	.5	.4

Table 2:

Example 2 *Let us consider the following graph.*

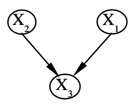

Figure 1:

We have: $(\Delta(x_1) = 0, \Pi(x_1) = .5)$,
$(\Delta(x_2) = 0, \Pi(x_2) = 1)$, $(\Delta(\neg x_1) = .45, \Pi(\neg x_1) = 1)$,
$(\Delta(\neg x_2) = .4, \Pi(\neg x_2) = .5)$.
Table 1 gives the conditional possibilities $(\Delta(X_3 \mid X_1X_2), \Pi(X_3 \mid X_1X_2))$.
Table 2 shows the possibility distributions π^* *and* π_* *associated with the graph.*

It is easy to check that $\forall u, \pi_*(u) \leq \pi^*(u)$.
However, the coherence property is not always guaranteed. Indeed, consider a bipolar possibilistic graph composed of two distinct nodes X and Y, such that:
$(\Delta(x), \Pi(x)) = (.3, .3)$, $(\Delta(y), \Pi(y)) = (.5, .6)$
$(\Delta(\neg x), \Pi(\neg x)) = (0, 1)$, $(\Delta(\neg y), \Pi(\neg y)) = (0, 1)$.
We have $\pi^*(xy) = .3$ while $\pi_*(xy) = .5$.

One way to guarantee that $\pi_* \leq \pi^*$ is to replace (14) by:

$$\pi_*(u) = min(min\{\Pi(x \mid u_{Par(X)}) : u \models x \text{ and } u \models u_{Par(X)}\},$$
$$max\{\Delta(x \mid u_{Par(X)}) : u \models x \text{ and } u \models u_{Par(X)}\}).$$

This expresses a revision of the lower distribution by means of the upper distributions in order to restore consistency.

7 Reasoning with fuzzy rules

7.1 What is a Rule ?

"If... then..." rules have been widely used for knowledge representation purposes in Artificial Intelligence since the late sixties, with the emergence of expert systems, and in Fuzzy Logic, with the widespread popularity of fuzzy control and fuzzy systems, since the mid seventies. To-date, rules, also called association rules, are found to be a convenient format for describing interpretable knowledge extracted from data. However, there has been little consensus on the exact nature of rules and their correct mathematical model. In the tradition of logic, as inherited by logic programming, a rule is modeled by a material implication. However, the fact that, when doing so, a rule takes the truth-value "true" when the condition part is false has led to paradoxes in logic, and to computational and representational difficulties in logic programming. In the latter area, researchers have felt the need to change the semantics of negation, so as to circumvent these difficulties (negation by failure, which is a form of CWA). In the tradition of expert systems, a rule is understood as a production rule, associated to a modus-ponens-like deduction process. A "rule" is thus a kind of inference rule, however without a clear mathematical status. In more recent probabilistic expert systems, rules are encoded as conditional probabilities in a belief network. This view of a weighted rule, if mathematically sound, is at odds with the logical tradition, since the probability of a material implication describing a rule clearly differs from the corresponding conditional probability. This observation [24] has led to a vivid debate in philosophical circles since the late seventies [20] without fully settling the case.

More recently some scholars like Calabrese [6], Goodman et al. [19], Dubois and Prade [13] have suggested that a rule is not a two-valued entity, but a three valued one. To see it, consider a database containing descriptions of items in a set S. If a rule "if A then B" is to be evaluated in the face of this database, it clearly creates a 3-partition of S, namely:

- the set of examples of the rule: $A \cap B$ [1],
- the set of counter-examples to the rule: $A \cap \overline{B}$,

[1] When A and B are not defined on the same universe, the cylindrical extensions are assumed. Then, for the sake of simplicity, $A \cap B$ is given for $(A \times V) \cap (U \times B)$, and $A \cup B$ is given for $(A \times V) \cup (U \times B)$.

- the set of items for which the rule is irrelevant: $\overline{A}$.

Each case should be encoded by means of a different truth-value. The two first cases only corresponding to the usual truth-values "true" and "false" respectively. The third case corresponds to a third truth-value which, according to the context, can be interpreted as "unknown", undetermined, irrelevant, etc. This idea of a rule as a "tri-event" actually goes back to De Finetti (1936) [7] who even proposed it as a genuine example of three-valued proposition which could be captured by Lukasiewicz logic. It is also the backbone of De Finetti's approach to conditional probability [8]. Indeed it is obvious to see that the probability $P(B|A)$ is entirely defined by $P(B \cap A)$ and $P(A \cap \overline{B})$. More recently it was shown by Dubois and Prade [13] that the three-valued semantics of rules was a possible semantics for the calculus of conditional assertions of Kraus, Lehmann and Magidor [23], which is the basis of one approach to nonmotonic reasoning. It is also in full accordance with Adams conditional logics [1] based on infinitesimal probabilities.

This framework for modeling a rule produces a precise mathematical model: a rule is modeled as a pair of disjoint sets representing the examples and the counter-examples of a rule, namely $(A \cap B,\ A \cap \overline{B})$. This definition has several consequences. First, it justifies the claim made by De Finetti that a conditional probability is the probability of a particular entity which can be called a conditional event.

Moreover it precisely shows that material implication only partially captures the intended meaning of an "if... then..." rule. It is obvious that the set of items where the material implication $\overline{A} \cup B$ is true is the complement of the set of counter-examples of a rule. Hence the usual logical view does not single out the examples of the rule, only its counter-examples. This is clearly in agreement with the fact that propositions in classical logic represent negative information. On the other hand, the set of examples of a rule is $A \cap B$ and clearly represents positive information.

Note that in the data mining literature, the merit of an association rule $A \Longrightarrow B$ extracted from a database is evaluated by two indices: the support degree and the confidence degree, respectively corresponding to the probability of the conjunction and the conditional probability. This proposal may sound ad hoc. However the deep reason why two indices are necessary to evaluate the quality of a rule is because the rule generates a 3-partition of the database, and two evaluations are needed to picture their relative importance. In fact the primitive quality indices of an association rule are the proportion of its examples and the proportion of its counter-examples. All other indices derive from these basic evaluations.

The three-valued representation of an "if... then..." rule also strongly suggests that a rule contains both positive and negative information embedded in it. It also sheds light on the debate between Mamdani's rules in fuzzy logic versus implicative rules: they are the two sides of the notion of a rule. It makes no sense to discard one of them. However each side may serve a specific purpose.

For instance, the three-valued representation of a rule may solve some well-known paradoxes of confirmation discussed by scholars casting new light on a question which goes back to Nicod and Hempel [21]. Consider the statement "all ravens are black" modeled in first order logic as:

$$\forall x,\ \text{Raven}(x) \rightarrow \text{Black}(x)\ \ , \tag{15}$$

whose validity is to be experimentally tested. Observing black ravens indeed confirms this statement while the observation of a single non-black one would refute it. However since this sentence also writes $\forall x,\ \neg\text{Raven}(x) \vee \text{Black}(x)$, it is logically equivalent to its contraposed form:

$$\forall x, \neg\text{Black}(x) \rightarrow \neg\text{Raven}(x)\ \ . \tag{16}$$

So observing a white swan, which does not violate the sentence, should also confirm the claim that all ravens are black. This is hardly defendable. Using a three-valued representation of the rule, it is clear that while black ravens are examples of the rule "all ravens are black", white swans are irrelevant items for this rule. Conversely, white swans are examples of the rule "all non-black items are non-ravens", and black ravens are in turn irrelevant. Both rules have the same set of counterexamples (all ravens that are not black, if any) but their examples differ, and only examples should meaningfully confirm rules. Hence the three-valued representation enables a rule to be distinguished from its contraposed form.

This representation also tolerates non-monotonicity, as claimed above. It is intuitively satisfying to consider that that a rule $R_1 =$ "if A then B" is safer than a rule $R_2 =$ "if C then D", if R_1 has more examples and less counterexamples than R_2 (in the sense of inclusion). Then R_1 is said to be entailed by R_2 (see Goodman et al. [19]). In particular, the two inclusions $C \cap D\ \subseteq\ A \cap B$ and $A \cap \overline{B}\ \subseteq\ C \cap \overline{D}$ generally ensure that $P(B \mid A)\ \geq\ P(D \mid C)$.

When rules are represented by material implications, it is easy to see that R_2 classically entails R_1 as soon as R_1 has less counterexamples than R_2, since it means $\overline{C} \cup D\ \subseteq\ \overline{A} \cup B$. Examples of the rules are never involved in classical entailment. As a consequence "if A then B" always implies "if A and C then B", classically. However using the three-valued representation, this is no longer the case. Indeed while the rule "if A and C then B" has generally less counterexamples than the rule "if A then B", it also has less examples, since $A \cap B \cap C\ \subseteq\ A \cap B$. So the rule "if A and C then B" cannot be a consequence of the rule "if A then B" any more in the three-valued setting.

Now, consider a rule "if X is A then Y is B", where X and Y are two attributes with domains U and V respectively. The sets A and B are now viewed as subsets of attribute values in U and V respectively. This rule implicitly defines a relation R between these domains. However the outline of this relation, as prescribed by the rule, depends on the way it is interpreted:

1. **Negative view.** The rule is viewed as a constraint of the form "if X is A, then Y *must* be B". In other words, if $x \in A$ and $y \notin B$, then $(x, y) \notin R$, or equivalently $R \subseteq \overline{A} \cup B$. This view emphasizes only the counter-examples to the rule. It is the implicative form of the rule. Pairs of attribute values in $A \cap \overline{B}$ are deemed impossible. Other ones remain possible.

2. **Positive view.** The rule is viewed as a case of the form "if X is A, then Y *can* be B". In other words, if $x \in A$ and $y \in B$, then $(x, y) \in R$, or equivalently $A \cap B \subseteq R$. This view emphasizes only the examples of the rule. It is the conjunctive form of the rule. Pairs of attribute values in $A \cap B$ are guaranteed possible. It is not known if other ones are possible or not.

The full picture of a rule of the form "if X is A then Y is B" involves these two sides and corresponds to a double inclusion:

$$A \cap B \ \subseteq \ R \ \subseteq \ \overline{A} \cup B \ . \qquad (17)$$

Namely the rule specifies a subset of the Cartesian Product $U \times V$ which includes examples and excludes counter-examples.

This dual view of fuzzy rules, and more particularly the advantages of using conjointly implicative rules (encoding negative information) and conjunctive rules (encoding positive information) in the same rule-based system, has been emphasized in [28].

This basic distinction between positive and negative information has never been emphasized in classical rule-based systems (e.g. expert systems). However, this is not really surprising. Generally, in these systems, when the rules encode a mapping from a numerical domain U to another one V, the condition parts of the rules are defined from a classical partition of U. As they do not overlap, a precise input triggers only one rule, which gives only one conclusion on V. Then it does not matter whether the conclusion encodes positive or negative information. By contrast, when dealing with uncertainty (imprecision, fuzziness, etc.) the condition parts of rules may overlap, and the inputs may be imprecise or fuzzy. Then, several rules may be triggered, each giving a (different) conclusion on the same domain. Hence these conclusions have to be aggregated, and their nature (positive or negative) determines the way they have to be aggregated, as shown in the following. It can explain why this distinction between positive and negative information first appeared in fuzzy rule-based systems. However, it is also relevant in some classical systems, where it could clarify the aggregation processes. For instance some data mining algorithms produce rules whose condition and conclusion parts are not subject to any a priori constraint. As the condition parts of these rules may overlap, a precise input can trigger several rules, whose conclusions have to be aggregated. Usually, this aggregation is ad-hoc, or based on some evaluation function of the produced outputs. Sometimes, only one rule is chosen as the most "representative" one. As a consequence, these methods do not take into account the semantics nor the potential incoherence of the produced rules.

7.2 Retrieving the Typology of Fuzzy Rules

Let $A \circ R$ denote the upper image of A through R:

$$A \circ R \ = \ \{v \,/\, \exists u \in A,\ (u, v) \in R\} \ .$$

It is interesting to notice the following equivalences.

- For implicative rules, $R \ \subseteq \ \overline{A} \cup B$ is equivalent to $A \circ R \ \subseteq \ B$. The latter inclusion is a formal translation of the linguistic expression corresponding to the rule: "if X is A, then Y *must* be B".
- For conjunctive rules, $A \cap B \ \subseteq \ R$ is equivalent to $A \circ \overline{R} \ \subseteq \ \overline{B}$. The latter inclusion is a bit less obvious. It means that if a value v for Y is not reachable from a value u for X in A, then (u, v) cannot be an example of the rule.

These four inclusions lead to four approaches to the modeling of fuzzy rules by fuzzy relations. Indeed, when considering fuzzy rules, these equivalences no longer hold. It means that, regardless of the choice of particular fuzzy connectives for conjunction or disjunction, four kinds of fuzzy rules fundamentally exist, as pointed out in [14].

7.2.1 Implicative Rules

Consider a knowledge base $\mathcal{K} = \{A_i \to B_i,\ i = 1, \ldots, n\}$, made of n parallel fuzzy rules (i.e., rules with the same input space U and the same output space V). These rules are interpreted as constraints (negative information). The fuzzy connective representing the arrow must thus be a generalized material implication. Each rule "if X is A_i, then Y *must* be B_i" (denoted $A_i \to B_i$) is represented by a conditional possibility distribution $\pi^i_{Y|X} = \mu_{A_i \to B_i}$ (the membership function of $A_i \to B_i$), which is determined according to the semantics of the rule. X is the tuple of input variables (on which information can be obtained) and Y is the tuple of output variables (about which we try to deduce information). If the implication is such that $1 \to a = a$, for $X = u \in A$ where A is not fuzzy, the fuzzy conclusion of a fuzzy rule is a fuzzy constraint on the values of Y. Each output value v is at most $\mu_B(v)$-possible:

$$\pi(u, v) \leq \mu_B(v), \quad \text{for all} \quad v \in V \ .$$

According to (18), the possibility distribution $\pi^{\mathcal{K}}$ representing the base $\mathcal{K}$ is obtained as the (min-based) conjunction of the $\pi^i_{Y|X}$'s:

$$\pi^{\mathcal{K}} = \min_{i=1,\ldots,n} \pi^i_{Y|X} \quad . \tag{18}$$

This equation shows that such rules are viewed as (fuzzy) constraints since the more rules, the more constraints, the smaller the number of values that satisfy them, and the smaller the levels of possibility. The possibility distribution $\pi^{\mathcal{K}}$ is then an upper bound of possibility for pairs of values in $U \times V$. Moreover, this conjunctive combination implies that some output values, which are possible according to some rules, can be forbidden by other ones. Then, the possibility degree $\pi^{\mathcal{K}}(u, v) = 0$ means that if $X = u$, then v is an impossible value for Y, i.e., (u, v) is an impossible pair of input/output values. By contrast, $\pi^{\mathcal{K}}(u, v) = 1$ denotes ignorance only. It means that for the input value $X = u$, no rule in $\mathcal{K}$ forbids the value v for the output variable Y. However, the addition of a new rule to $\mathcal{K}$ (expressing a new piece of knowledge) may lead to forbid this value. A possibility degree $\pi^{\mathcal{K}(\sqcap,\sqsubseteq)} > 0$ means that the pair (u, v) is not known as totally impossible, with respect to the current knowledge.

Note that since, with the negative information view, rules express constraints on possible situations, it may happen that a knowledge base is inconsistent with a given input value $X = u$. In the fuzzy case, it means that the fuzzy set of possible values $\{(v, \pi(u, v)),\ v \in V\}$ is sub-normalized, if not empty. Then the rule base is partially useless as inference cannot be carried out for some input values. This feature has been considered an impediment when using implicative rules for fuzzy modeling. Indeed using the Takagi-Sugeno approach or the Mamdani approach, this situation never occurs. However, in the logic of negative information, it can be viewed as an expected and valuable feature. It means that this approach to fuzzy rule-based systems enables them to be tested for logical consistency [15], which can be used to force the human expert providing the rules (or the learning system) to propose meaningful sets of rules (forbidding rule bases containing statements like "if X is A, then Y is B" and "if X is A, then Y is *not* B" simultaneously, for instance). Generally, a set of implicative rules is logically coherent if and only if the fuzzy set of possible values $\{(v, \pi(u, v)), v \in V\}$ is normalized for all $u \in U$.

According to the typology of fuzzy rules proposed in [14], there are two main kinds of implicative rules, whose prototypes are *certainty* and *gradual* rules.

Certainty rules

are of the form "The more X is A, the more certainly Y lies in B", as in "The younger a man, the more certainly he is single". They are based on the fuzzy translation of the inclusion $R \subseteq \overline{A} \cup B$. They correspond to the following conditional possibility distribution modeling the rule:

$$\forall (u, v),\ \pi_{Y|X}(v, u) \leq \max(\mu_B(v), 1 - \mu_A(u)) \quad . \tag{19}$$

In (19), A and B are combined with Kleene-Dienes implication: $a \to b = \max(1-a, b)$. Given such a rule, and an observation A' for the input variable X, the (fuzzy) set B' denotes the induced restriction on the possible values for the output variable Y. For a precise input $A' = \{u_0\}$, the conclusion B' is such that $\forall v \in V,\ \mu_{B'}(v) \geq 1 - \mu_A(u_0)$, i.e., a uniform level of uncertainty $1 - \mu_A(u_0)$ appears in B'. Then, "Y is B" is certain only to the degree $\mu_A(u_0)$, since values outside B are possible to the complementary degree. A similar behavior is obtained with the implication $a \to b = 1 - a \star (1-b)$, where $\star$ is the product instead of min. Clearly, certainty rules extend propositions with certainty levels by making the certainty level depend on the level of satisfaction of fuzzy properties.

Gradual rules

are of the form "The more X is A, the more Y is B", as in "The redder the tomato, the riper it is". They are based on the fuzzy translation of the inclusion $A \circ R \subseteq B$. They correspond to the constraint:

$$\forall u \in U,\ \mu_A(u) \star \pi_{Y|X}(v, u) \leq \mu_B(v)\ , \tag{20}$$

where $\star$ is a conjunction operation. The greatest solution for $\pi_{Y|X}(v, u)$ in (20) (according to the minimal specificity principle which calls for the greatest permitted degrees of possibility) corresponds to the residuated implication:

$$\mu_{A \to B}(u, v) = \sup\{\beta \in [0, 1],\ \mu_A(u) \star \beta \leq \mu_B(v)\}\ . \tag{21}$$

When $\star$ is min, (21) corresponds to Gdel implication: $a \to b = 1$ if $a \leq b$, and b if $a > b$. If only a crisp relation between X and Y is supposed to underlie the rule, it can be modeled by Rescher-Gaines implication: $a \to b = 1$ if $a \leq b$, and 0 if $a > b$.

For an input u_0, (21) provides an enlargement of the core of B, i.e., the less X satisfies A, the larger the set of values near the core of B which are completely possible for Y. This expresses a similarity-based tolerance: if the value of X is close to the core of A, then Y is close to the core of B and, more precisely, the closer X is to the core of A, the closer Y must be to the core of B. This is the basis for an interpolation mechanism when dealing with several overlapping rules ([9],[16]).

7.2.2 Conjunctive Rules

In the fuzzy control tradition, rule-based systems are often made of conjunction-based rules, as Mamdani-rules for instance. ¿From the previous section it is clear that this representation considers fuzzy rules as cases, pieces of positive information. These cases, denoted $A_i \wedge B_i$ represent clusters of imprecise observations, i.e., couples of jointly possible input/output values. Each conjunctive rule is then represented by a joint "guaranteed possibility" distribution $\delta^i_{X,Y} = \mu_{A_i \wedge B_i}$.

If the conjunction $\wedge$ is such that $1 \wedge a = a$, for a precise input $X = u \in A$ where A is not fuzzy, the fuzzy conclusion of a fuzzy rule $A \wedge B$ is a collection of examples of feasible or reachable values for Y. Namely, each value $v \in V$ is at least $\mu_B(v)$-possible. If the rule does not apply (i.e., if $X = u \notin A$), then the inference mechanism leads to the conclusion $Y = \emptyset$ which, in the context of positive information means that Y is unknown, in the sense that no output is guaranteed possible.

Since conjunctive rules model imprecise information, the guaranteed possibility distribution representing the database is obtained by a disjunctive combination of the conjunctive rules, which appropriately corresponds to the accumulation of observations and leads to those pairs of values (u, v) whose possibility/feasibility is guaranteed to some minimal degree. Then, the counterpart to (18) is:

$$\delta_{\mathcal{K}} = \max\nolimits_{i=1,\ldots,n} \delta^i_{X,Y} \quad . \tag{22}$$

The distribution $\delta_{\mathcal{K}}$ is then a lower bound of possibility for pairs of values in $U \times V$.

Thus, a possibility degree $\delta_{\mathcal{K}}(u, v) = 1$ means that if $X = u$, then v is a totally possible value for Y. This is a guaranteed possibility degree. By contrast, $\delta_{\mathcal{K}}(u, v) = 0$ only means that if $X = u$, no rule can guarantee that v is a possible value for Y. By default, v is considered as not possible (since possibility cannot be guaranteed). A membership degree 0 to B' represents ignorance, while a degree 1 means a guaranteed possibility. Thus, a conclusion $B' = \emptyset$ should not be understood here as "all the output values are impossible", but as "no output value can be guaranteed".

Among objections to fuzzy rule aggregation a la Mamdani coming from the logical camp is the observation that adding an empty rule "if anything, then anything", to a rule base destroys the available information (because $(\bigcup_i A_i \times B_i) \cup (U \times V) = U \times V$). This objection would be valid in the negative information view where the rule "if anything, then anything" has indeed empty contents (it is a tautology), and the disjunctive combination of information is absurd. However, in the logic of positive information, the rule "if X is U then Y is V" is not a tautology but has very extreme contents. It says that whatever the (input) value u for X, any value v for the output variable Y is guaranteed possible. In other words, there is absolutely no relationship between X and Y. No surprise if this type of positive information should destroy all previous observations. With this model, the empty rule is obviously "If X is U, then Y is $\emptyset$" which means that whatever $u \in U$, no value can be guaranteed possible for Y. Then, suitably, $(\bigcup_i A_i \times B_i) \cup (U \times \emptyset) = \bigcup_i A_i \times B_i$.

As for implicative rules, there are two main kinds of conjunctive rules, called *possibility* and *antigradual* rules [14].

Possibility rules

are of the form "the more X is A, the more possible Y lies in B", as in "the more cloudy the sky, the more possible it will rain soon". They are based on the inclusion $A \cap B \subseteq R$. They correspond to the following possibility distribution modeling the rule:

$$\forall (u, v) \in U \times V,\ \delta_{X,Y}(u, v) \geq \min(\mu_A(u), \mu_B(v)) \ . \tag{23}$$

These rules, modeled with the conjunction *min*, correspond to the ones introduced by Mamdani and Assilian in 1975 [25]. For an input value u_0 such that $\mu_A(u_0) = \alpha$, a possibility rule expresses that, when $\alpha = 1$, B is a set of possible values for Y (to different degrees if B is fuzzy). When $\alpha < 1$, values in B are still possible, but they are guaranteed possible only up to the degree α. To obtain B', the output set associated with u_0, the set B is then truncated from above. Finally, if $\alpha = 0$, the rule does not apply, and $B' = \emptyset$ as already said.

Antigradual rules

which are another type of conjunctive rule (see [14]), express that "the more X is A, the larger the set of guaranteed possible values for Y is, around the core of B", as in "the more experienced a manager, the wider the set of situations he can manage". They are based on the inclusion $A \circ \overline{R} \subseteq \overline{B}$. For an input $A' = \{u_0\}$, if $\mu_A(u_0) = \alpha < 1$, the values in B such that $\mu_B(v) < \alpha$, cannot be guaranteed. Such a rule expresses how values which are guaranteed possible can be extrapolated on a closeness basis.

8 Bipolar preferences

Bipolar possibilistic logic provides a natural framework for representing positive and negative desires, as recently advocated in [4]. We only summarize the main ideas here. Indeed, negative desires which refer to what is more or less rejected by the considered agent, correspond to constraints of the form $\Pi(R_k) \leq \rho_k$. They express that the models of R_k are considered as rather impossible, since their level of possibility, of satisfactoriness are upper bounded by ρ_k. By duality, we obtain the equivalent constraint that $N(R_k^c) \geq 1 - \rho_k$ which emphasize that the models, in R_k^c, of $\neg r_k$ are not rejected and thus acceptable with certainty $1 - \rho_k$. It can be represented by the possibility distribution $\pi_k^*(u) = 1$ if $u \models \neg r_k$ and $\pi_k^*(u) = \rho_k$ if $u \models r_k$. Then a conjunctive aggregation of these k elementary distributions, in the sense of Definition 1, provides the fuzzy set, with membership function $\pi^*(u) = min_k \pi_k^*$, of interpretations u which are not rejected.
Besides positive desires are represented by constraints of the form $\Delta(D_j) \geq \delta_j$, which expresses that any model of d_j is satisfactory at least at level δ_j. This is

represented by the possibility distribution $\pi_{j*}(u) = \delta_j$ if $u \models d_j$ and $\pi_{j*}(u) = 0$ if $u \models \neg d_j$. The positive desires $[d_j, \delta_j]$ are thus aggregated disjunctively, i.e. $\pi_*(u) = max_j \pi_j(u)$, in agreement with Definition 2.
This representation framework has been applied in [4] for fusing the positive and negative preferences of different agents. Even if the preferences of each agent t are consistent in the sense that $\forall u, \pi_*^t \leq \pi^{t*}$, it may occur, that the result of the fusion of the lower distributions is no longer consistent with the result of the fusion of the upper distributions, indeed the same aggregation operator is not used for fusing the positive preferences represented by π_*^t on the one hand, and the non-rejection-based preferences encoded by π^{t*} on the other hand. Then a revision step should take place for enforcing the consistency. The result of the fusion of the positive desires is revised if necessary for preserving consistency with what is not rejected by any of the agents [4]. The fusion operations as well as the revision step can be directly performed at the syntactic level on the bases $\{(\neg r_k, 1 - \rho_k)\}_k$ and $\{[d_j, \delta_j]\}_j$, see [4].

9 Conclusion

This paper has presented and advocated a twofold framework for representing uncertainty qualitatively. This can be useful for knowledge representation as well as preference representation. The bipolarity enables us to distinguish between what is possible or satisfactory for sure from what is just not impossible, or not undesirable.
Bipolarity takes place each time that we have to specify a *subset* of values about which we have incomplete or uncertain information. This situation is frequently encountered, especially when dealing with generic knowledge, or with preferences.

Besides, a possibility distribution can be represented equivalently in various compact frameworks: as a possibilistic logic base, as a possibilistic graph, or as strict comparative possibility constraints. This has been extended to the bipolar setting, by first defining the idea of conditioning (underlying the two last compact representation frameworks listed above) for guaranteed possibility measures. Algorithms exist for going directly from one compact representation to another [3]. They can be adapted to the bipolar framework. Efficient algorithms recently developed for uncertainty propagation in possibilistic graphs [2] could be extended to bipolar possibilistic graphs as well.
Lastly, the bilattice structure [18] of the bipolar framework is worth investigating and may be of interest, especially in the view of developing possibilistic logic programming, following Fitting [17]'s approach.

References

[1] E.W. Adams. *The logic of conditionals.* D. Reidel, Dordrecht, The Netherlands, 1975.

[2] N. Ben Amor, S. Benferhat, and K. Mellouli. Anytime possibilistic propagation algorithm. In *Proceedings of the 1st International Conference on Computing in an Imperfect World, Soft-Ware 2002,Springer Verlag LNCS 2311*, pages 263–279, 2002.

[3] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bridging logical, comparative and graphical possibilistic representation frameworks. In *Proceedings of the 6th Eur. Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'01)*, pages 422–431, 2001.

[4] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar representation and fusion of preferences in the possibilistic logic framework. In *Proceedings of the eighth International Confenrence on Principle of Knowledge Representation and Reasoning (KR'02)*, pages 158–169, 2002.

[5] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In *Proceedings of 3rd International Conference of Principles of Knowledge Representation and Reasoning (KR'92)*, pages 673–684, 1992.

[6] P. Calabrese. An algebraic synthesis of the foundations of logic and probability. *Information Sciences*, 42:187–237, 1987.

[7] B. De Finetti. La logique des probabilits. In *Actes du Congrs International de Philosophie Scientifique*, volume IV, pages 1–9, Paris, France, 1936. Hermann et Cie, Paris, France.

[8] B. De Finetti. *Theory of Probability — A critical Introductory Treatment*, volume 1 & 2. John Wiley & Sons, Chichester, United Kingdom, 1974.

[9] D. Dubois, M. Grabisch, and H. Prade. Gradual rules and the approximation of control laws. In H.T. Nguyen, M. Sugeno, R. Tong, and R.R. Yager, editors, *Theoretical Aspects of Fuzzy Control*, pages 147–181. Wiley, 1994.

[10] D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *Journal of Logic, Language, and Information*, 9:65–89, 2000.

[11] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. *In Handbook of Logic in Artificial Intelligence and Logic Programming*, (D. Gabbay et al., eds, 3, Oxford University Press:pages 439–513, 1994.

[12] D. Dubois and H. Prade. (with the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale). Possibility Theory - an Approach to Computerized Processing of Uncertainty. 1988.

[13] D. Dubois and H. Prade. Conditional objects as non-monotonic consequence relationships. *IEEE Transactions on Systems, Man and Cybernetics*, 24(12):1724–1739, 1994.

[14] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84(2):169–186, 1996. Special issue in memory of Prof A. Kaufmann.

[15] D. Dubois, H. Prade, and L. Ughetto. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Transactions on Fuzzy Systems*, 5(3):398–417, 1997.

[16] D. Dubois, H. Prade, and L. Ughetto. Fuzzy logic, control engineering and artificial intelligence. In H.B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control Engineering*, pages 17–57. Kluwer Academic Publishers, 1999.

[17] M. Fitting. Bilattices in logic programming. In *The 20th International Symposium on Multiple-Valued Logic*, pages 238–246. George Epstein, editor, 1990.

[18] M. L. Ginsberg. Multivalued logics: a uniform approach to reasoning in artificial intelligence. *Compuational Intelligence*, 4:265–316, 1988.

[19] I.R. Goodman, H.T. Nguyen, and E. Walker. *Conditional Inference and Logic for Intelligent Systems: a theory of measure-free conditionals*. North-Holland, Amsterdam, The Netherlands, 1991.

[20] W.L. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs*. D. Reidel, Dordrecht, The Netherlands, 1981.

[21] C.G. Hempel. Studies in the logic of confirmation. *In Aspects of scientific explanation and other essays in the philosophy of science*, pages 3–46, 1965.

[22] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, 1, 1978.

[23] K. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[24] D. Lewis. Probabilities of conditionals and conditional probabilities. *Journal of Philosophical Logic*, 3, 1973. Reprinted in Harper et al. [20].

[25] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal on Man-Machine Studies*, 7:1–13, 1975.

[26] S. Nahmias. Fuzzy variables. *Fuzzy Sets and Systems*, 1:97–110, 1978.

[27] Y. Shoham. Reasoning About Change. In *MIT Press, MA*, 1988.

[28] L. Ughetto, D. Dubois, and H. Prade. Implicative and conjunctive fuzzy rules - A tool for reasoning from knowledge and examples. In *Proceedings of the 16th American National Conference on Artificial Intelligence (AAAI'99)*, pages 214–219, Orlando, Fl, USA, 1999.

[29] L. van der Torre and E. Weydert. Parameters for utilitarian desires in a qualitative decision theory. *Applied Intelligence*, 14:285–301, 2001.

[30] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.