Fuzzy Sets in Approximate Reasoning: A Personal View

DIDIER DUBOIS and HENRI PRADE

ABSTRACT

Fuzzy set-based methods contribute to the formalization of different types of approximate reasoning, mainly in two ways. First, when modeling classes and properties, fuzzy sets naturally encode gradual properties, such as, e.g., 'large', whose satisfaction is a matter of degree and may be only partial for particular instances. Due to this ability, fuzzy sets are naturally entitled to capture intermediary situations and to be instrumental in the formalization of interpolative reasoning. More generally, similarity-based approximate reasoning can greatly benefit from fuzzy set approaches since similarity is usually a matter of degree. Second, fuzzy sets can also represent incomplete information pervaded with uncertainty. They are then viewed as possibility distributions and give birth to possibility and necessity measures to assess the degrees to which a statement is possible or is certain taking into account the available (incomplete) information. Fuzzy rules whose conclusion part are uncertain can then be modelled in this framework and used for deduction purposes. Moreover, possibility theory can handle default reasoning where plausible conclusions are drawn in a given state of information but may be later defeated when they are inconsistent with a new piece of information which becomes available. The possibilistic modelling of uncertainty is also shown to be useful in abductive reasoning for diagnosis problems. The paper provides an overview of the way fuzzy sets can serve for different types of reasoning purposes and gives the background necessary to the understanding of the basic methodological issues.

1 INTRODUCTION

Fuzzy rule-based approximate reasoning is raising more and more interest from researchers and practitioners nowadays. Zadeh ([Zad73], [Zad75], [Zad79a]) provided the basic machinery for fuzzy set-based approximate reasoning more than fifteen years ago. In his approach, each granule of knowledge is represented by a fuzzy set or a fuzzy relation on the appropriate universe. Then the fuzzy set representations of the different granules are combined and the result of this combination is projected on the universe(s) of interest. A well-known particular case of this method is the pattern of inference, named "generalized modus ponens" which enables us to deduce a fuzzy conclusion from a fuzzy rule and a fuzzy fact pertaining to the universe of discourse associated with the condition part of the rule. However, this framework is rather general, and a proper application of it requires a correct understanding of the intended meaning of the pieces of knowledge to be represented by fuzzy sets. In particular, fuzzy rules may have very different semantics, which lead to different choices concerning the multiple-valued connective to be used for modeling the rule. In this paper, we distinguish between purely gradual rules and rules with uncertain conclusion parts. Purely gradual rules are of the form "the more X is A, the more Y is B" which qualitatively describes a relation between the values of X and Y, but which is not pervaded with uncertainty. Gradual rules of the form "the closer X is to…, the closer Y is to…" express pieces of knowledge which are of interest for interpolative reasoning purposes. Patterns of reasoning under uncertainty involve rules of the forms "if p is true then q is somewhat certain", "the more X is A, the more certain Y is B", or "the more X is A, the more possible Y is B", which carry different kinds of uncertainty in their conclusion parts. We shall show in the following how these different kinds of rules can be represented in the fuzzy set framework.

In this paper we make a strong distinction between the use of fuzzy sets for modelling gradual properties, i.e., properties whose satisfaction is a matter of degree, or gradual notions like similarity, and the use of fuzzy sets for modelling incomplete, imprecise or uncertain information. In the latter case, the fuzzy set-based representation of the available information induces possibility and necessity measures for assessing the uncertainty of the conclusions which are computed. It is worth emphasizing that possibility theory offers an ordinal framework for the modeling of uncertainty due to the use in the theory of operations like max or min which are only sensitive to the ordering of the operands and of an order-reversing operation (the complementation to 1 when using the scale [0,1]).

The paper is organized in two main parts. The first one is devoted to the use of the ideas of graduality expressed in terms of fuzzy sets for interpolative and qualitative reasoning purposes. The second main part deals with the management of uncertainty, and more particularly with the modelling of uncertain rules, the treatment of uncertainty and inconsistency in possibilistic logic, a possibility measure-based approach to default reasoning, and a possibilistic treatment of incomplete information in abductive reasoning for diagnosis.

2 GRADUALITY- AND SIMILARITY-BASED APPROXIMATE REASONING

One of the natural semantics for fuzzy sets is the expression of closeness, proximity, similarity. Then the elements in the core (or peak) of a fuzzy set (i.e., the ones with membership 1) are viewed as prototypical elements of the fuzzy set, while the other membership grades estimate the closeness of the elements to the prototypical ones. Proximity or similarity viewed as a graded relation between two elements of a referential can be represented in terms of a binary fuzzy relation, where the closer to 1 the degree of membership of a pair of elements to the fuzzy relation, the greater their proximity or their similarity. In this section, we focus on interpolative reasoning based on a set of fuzzy rules associated with similarity relations expressing closeness and then we consider the use of approximate equality relations for qualitative reasoning on orders of magnitude. We first give the necessary background on fuzzy relations and related issues.

2.1 Composition of Fuzzy Relations, Extension Principle, and Similarity

A binary fuzzy relation R is defined as a fuzzy subset on a Cartesian product $U \times V$. The composition of classical relations, namely $S \circ R = \{(u,w), \exists v, (u,v) \in R$ and $(v,w) \in S\}$ is extended to fuzzy relations R and S defined on $U \times V$ and $V \times W$ respectively (we may have $U = V = W$) by

$$\forall\, u \in U, \forall\, w \in W, \mu_{S \circ R}(u,w) = \sup_{v \in V} \min(\mu_R(u,v), \mu_S(v,w)). \qquad (1)$$

More generally, min may be replaced by an operation $*$ and the $\circ_*$ composition is then defined as

$$\mu_{S \circ_* R}(u,w) = \sup_v \mu_R(u,v) * \mu_S(v,w) \qquad (2)$$

where $*$ is a binary operation defined on [0,1] expressing a conjunction such that $0 * 0 = 0 * 1 = 1 * 0 = 0$, $1 * 1 = 1$. The operation $*$ is usually chosen as being associative, monotonically increasing in each place and satisfying $a * 1 = a$, $\forall\, a \in [0,1]$. Such an operation is known as a triangular norm (see [SS83]) and can be shown to be commutative and such that $\forall\, (a,b) \in [0,1]^2$, $a * b \leq \min(a,b)$. Noticeable triangular norms are min, product and the linear conjunction $\max(0, a + b - 1)$. When $*$ is associative, the composition $\circ_*$ is also associative. We simply wrote $\circ$ if $* =$ min.

A particular case of (2) is when one of the fuzzy relations is replaced by a fuzzy set A (here on U), namely

$$\mu_{R \circ A}(v) = \sup_u \mu_R(u,v) * \mu_A(u). \qquad (3)$$

The composition $R \circ A$ extends to a fuzzy set A and to a fuzzy multiple-valued mapping Γ which associates with $u \in U$ the fuzzy set $\Gamma(u)$ defined by $\mu_{\Gamma(u)}(v) = \mu_R(u,v)$, the image of a classical subset by an ordinary multiple-valued mapping. A particular case of (3) is the extension principle ([Zad65], [Zad75]) which enables the image f(A) of a fuzzy set A via a function f from a set U to a set V to be computed as

$$\begin{aligned} \forall\, v \in V, \mu_{f(A)}(v) &= \sup_{u: v=f(u)} \mu_A(u) \\ &= 0 \text{ if } \nexists u, v = f(u). \end{aligned} \qquad (4)$$

When A is a crisp set, (1) reduces to the usual extension of a function to a set-valued argument, namely $f(A) = \{v, \exists\, u \in A, v = f(u)\}$. If f is bijective

$$\forall\, v \in V, \mu_{f(A)}(v) = \mu_A(f^{-1}(v)). \qquad (5)$$

When the function f has several arguments $u_1, u_2, \ldots, u_n$ an interesting situation is when the fuzzy set A is separable, so that A is a fuzzy Cartesian product $A_1 \times A_2 \times \ldots \times$

A_n with membership function $\mu_{A_1} * \mu_{A_2} * \ldots * \mu_{A_n}$ where $*$ is a triangular norm which models conjunctions in fuzzy set theory. Then the extension principle reads:

$$\mu_{f(A)}(v) = \sup\{\mu_{A_1}(u_1) * \mu_{A_2}(u_2) * \ldots * \mu_{A_n}(u_n) \mid f(u_1, u_2, \ldots, u_n) = v\} \qquad (6)$$

which is an interactive form of the extension principle (see, e.g., [DP87a]). In this paper, we will restrict to the non-interactive case, i.e., $*$ = min.

A fuzzy interval M is simply a fuzzy set of the real line such that $\mu_M(u) = 1$ for some u, and whose α-cuts $M_\alpha = \{u, \mu_M(u) \geq \alpha\}$ are closed intervals. The membership function of a fuzzy interval is thus upper semi-continuous (u.s.c.).The calculus of fuzzy intervals is based on the extension principle applied to real-valued real functions. Using u.s.c. fuzzy intervals M we are sure that any level cut $f(M)_\alpha$ of f(M) is equal to the image of the corresponding level cut M_α of M through f when f is continuous, i.e., $f(M)_\alpha = f(M_\alpha)$. Let the interval $(M_i)_\alpha$ be denoted $[m^-_i(\alpha), m^+_i(\alpha)]$. We have the following result (see, e.g., [DP87a]).

Let $M_1, \ldots, M_n$ be u.s.c. fuzzy intervals whose supports are bounded. Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuous, *non-decreasing* with respect to each of its arguments. Then, in the non-interactive case ($*$ = min), the image of $M_1, \ldots, M_n$ by f is an upper semi-continuous fuzzy interval whose α-cuts are:

$$[f(M_1, \ldots, M_n)]_\alpha = [f(m^-_1(\alpha), \ldots, m^-_n(\alpha)), f(m^+_1(\alpha), \ldots, m^+_n(\alpha))] \qquad (7)$$

for every α in [0,1]. We shall see in Section 2.3 that the extension of arithmetic operations to fuzzy intervals can be useful to compute the composition of fuzzy relations expressing approximate equalities for instance. For instance, taking $f(u_1,u_2) = u_1 + u_2$ in (7) yields for the extension of the addition to fuzzy intervals

$$(M_1 \oplus M_2)_\alpha = [m^-_1(\alpha) + m^-_2(\alpha), m^+_1(\alpha) + m^+_2(\alpha)] = (M_1)_\alpha \oplus (M_2)_\alpha.$$

Thus if M_i is a trapezoid with support (a_i,d_i) (i.e., the set of elements with non-zero membership values) and core$[b_i,c_i]$, then $M_1 \oplus M_2$ is a trapezoid with support$(a_1 + a_2, d_1 + d_2)$ and core$[b_1 + b_2, c_1 + c_2]$.

A similarity relation S is a binary fuzzy relation defined on a set U whose membership function satisfies the three following properties ([Zad71])

(i) $\forall\ u \in U,\ \mu_S(u,u) = 1$ (reflexivity)

(ii) $\forall\ u \in U,\ \forall\ v \in U,\ \mu_S(u,v) = \mu_S(v,u)$ (symmetry)

(iii) $\forall\ u \in U,\ \forall\ v \in U,\ \forall\ w \in U,\ \mu_S(u,v) * \mu_S(v,w) \leq \mu_S(u,w)$ (max-$*$ transitivity)

Fuzzy relations satisfying (i) and (ii) are called proximity relations. See ([Ovc91]) for a recent overview on similarity relations.

Similarity relations are closely related to the idea of distance. In particular if $*$ is the linear conjunction, max-$*$ transitivity is equivalent to the triangle inequality and $1 - \mu_S$ is a pseudometric; max-min transitivity, which is stronger, corresponds to ultrametrics. Max-$*$ transitivity can be also easily interpreted in terms of composition of fuzzy relations, namely (iii) can be written $S \circ_* S \subseteq S$ where fuzzy set inclusion is pointwisely defined by an inequality between membership functions. As we can see (iii) extends transitivity to degrees of similarity which are intermediary between 0 and 1 (transitivity of non-fuzzy relations is retrieved as a particular case). We shall see in Section 2.3 *weaker* forms of transitivity-like properties where the similarity or approximate equality slightly deteriorates when applying transitivity.

2.2 Interpolative Reasoning

Many authors have pointed out that fuzzy control techniques perform an interpolation between the conclusions of the fuzzy control rules, on the basis of the degrees of matching of the situation under concern with the condition parts of the different rules. Interpolative reasoning naturally subsumes some idea of proximity since intermediary values are computed from the known values corresponding to the situations which are the closest ones to the current one. So, it is not surprizing that the two main fuzzy control methods, namely Mamdani ([Mam77])'s and Sugeno ([Sug85])'s, can receive some justifications, starting with a proper handling of proximity relations.

Let us consider a collection of n fuzzy rules if X_1 is ${A^{(i)}}_1$ and… X_p is ${A^{(i)}}_p$ then Y is $B^{(i)}$, i = 1,n. Then Klawonn and Kruse ([KK93a],[KK93b]) have shown that if the (fuzzy) relation R relating the X_j's and Y satisfies some extensionality axioms, this is enough for deriving the result given by Mamdani's method before defuzzification. Let S and T be two similarity relations equipping the domains U and V of X and Y respectively. Then the extensionality of a relation R on $U \times V$ means

(i) $\forall\, u \in U, \forall\, s \in V, \forall\, v \in U, \mu_R(u,s) * \mu_S(u,v) \le \mu_R(v,s)$

(ii) $\forall\, u \in U, \forall\, s \in V, \forall\, t \in V, \mu_R(u,s) * \mu_T(s,t) \le \mu_R(u,t)$.

It expresses that if u and s are in relation R, u (resp.: s) should be also in relation with the neighbours of s (resp.: u). Then let us assume that the $A^{(i)}{}_j$'s and $B^{(i)}$ can be viewed as fuzzy sets of values close to peak values (here taken as unique for simplicity), i.e., $A^{(i)}{}_j = \{a^{(i)}{}_j\} \circ S_j$, $B^{(i)} = \{b^{(i)}\} \circ T$ where the $a^{(i)}{}_j$'s and $b^{(i)}$'s are peak values, S_j and T are similarity relations. Then, using the extensionality of the relation R relating the X_j's and Y, as well as the fact that $\mu_R(a^{(i)}{}_1, \ldots, a^{(i)}{}_p, b^{(i)}) = 1$, $\forall\, i = 1,n$ (the peak values define points which completely belong to the graph of R), and defining the similarity relation on $U_1 \times \ldots U_p$ as $\mu_{S_1} * \ldots * \mu_{S_p}$, leads to

$$\mu_R(x_1, \ldots, x_p, v) \ge \max_{i=1,n} \mu_{A^{(i)}{}_1}(x_1) * \ldots * \mu_{A^{(i)}{}_p}(x_p) * \mu_{B^{(i)}}(v). \tag{8}$$

When $*$ = min, we recognize Mamdani's expression of the fuzzy set of V to be defuzzified for computing an output Y = y when the observed input is $X_1 = x_1, \ldots, X_p = x_p$. However, at this point, no interpolation has been modelled in (8).

Sugeno's method starts from rules with non-fuzzy conclusion parts of the form "if X is $A^{(i)}$ then $Y = b^{(i)}(x)$ (we do not deal here with compound conditions for simplicity), and computes the output y as

$$y = \Sigma_i\, \mu_{A^{(i)}}(x) \cdot b^{(i)}(x) / \Sigma_i\, \mu_{A^{(i)}}(x) \tag{9}$$

which indeed looks like an interpolation. When $b^{(i)}(x)$ does not depend on x, this result can be obtained by applying Zadeh's approximate reasoning combination and projection approach viewing the rules as *gradual rules* expressing that "the closer X is to $a^{(i)}$, the closer Y is to $b^{(i)}$", i.e., modelling them as inequality constraints of the form $\mu_{A^{(i)}}(u) \le \mu_{B^{(i)}}(v)$. Then the subset of V obtained by combining the results of the rules for the input $X = x_0$ is given by $\min_{i=1,n} \mu_{A^{(i)}}(x_0) \rightarrow \mu_{B^{(i)}}(v)$ where the implication defined by $a \to b = 1$ if $a \leq b$ and $a \to b = 0$ if $a > b$ encodes the above interpretation of the rules. When the $A^{(i)}$'s and $B^{(i)}$'s make suitable fuzzy partitions of U and V respectively (it guarantees that $\forall u, \sum_i \mu_{A(i)}(u) = 1$), which is in particular the case with the usual partitions made of triangular membership functions as in Figure 1, it can be shown that the fuzzy subset of V which is thus obtained is nothing but the singleton $\{y\}$ computed by Sugeno's method (9). Let us explain the situation in more details.

Let us consider a collection of gradual rules of the form "the closer X is to a_i, the closer Y is to b_i" where (a_i,b_i), $i = 1,n$ are pairs of *scalar* values. The first problem is to represent "close to a_i", by means of a fuzzy set A_i. It seems natural to assume that $\mu_{A_i}(a_{i-1}) = \mu_{A_i}(a_{i+1}) = 0$ since there are special rules adapted to the cases $X = a_{i-1}$, $X = a_{i+1}$. Moreover if $u \neq a_i$, $\mu_{A_i}(u) < 1$ for $u \in (a_{i-1},a_{i+1})$, since information is only available for $u = a_i$. Hence A_i should be a fuzzy interval with support (a_{i-1},a_{i+1}) and core $\{a_i\}$. Besides, by symmetry, since the closer x is to a_{i-1}, the farther it is from a_i, $\mu_{A_{i-1}}$ should decrease when μ_{A_i} increases, and $\mu_{A_i}\left(\frac{a_i + a_{i+1}}{2}\right) = \mu_{A_{i-1}}\left(\frac{a_{i-1} + a_i}{2}\right) = 0.5$. The simplest way of achieving this is to let

$$\forall\, u \in [a_{i-1},a_i],\ \mu_{A_{i-1}}(u) + \mu_{A_i}(u) = 1$$

an example of which are triangular-shaped fuzzy sets as in Figure 1. Clearly the conclusion parts of the rules should involve fuzzy sets B_i whose meaning is "close to b_i", with similar convention. In other words, each rule is understood as "the more X is A_i, the more Y is B_i", i.e. here "the closer X is to a_i, the closer Y is to b_i". In that case the output associated with the precise input $X = a$, where $a_{i-1} < a < a_i$ is

$$B = (\alpha_{i-1} \to B_{i-1}) \cap (\alpha_i \to B_i) = (B_{i-1})_{\alpha_{i-1}} \cap (B_i)_{\alpha_i}$$

where $\alpha \to B$ corresponds to the level cut B_α, where $\to$ is the implication defined above and $\alpha_{i-1} = \mu_{A_{i-1}}(a)$, $\alpha_i = \mu_{A_i}(a)$ with $\alpha_{i-1} + \alpha_i = 1$. Then it can be easily proved (without the assumption of a triangular shape) that there exists a unique value $y = b$ such that $\mu_B(b) = 1$, which exactly corresponds to the result of a linear interpolation, i.e., we have $b = \alpha_{i-1} \cdot b_{i-1} + \alpha_i \cdot b_i$. It is a theoretical justification for Sugeno and Nishida ([SN85])'s inference method where the conclusion parts of the rules are precise values b_i and where a linear interpolation is performed on the basis of the degrees of matching $\alpha_i = \mu_{A_i}(a)$. Hence reasoning with gradual rules does model interpolation, linear interpolation being retrieved as a particular case. The more complicated case of gradual rules with compound conditions, i.e., rules of the form "the more X_1 is A_1, …, and the more X_n is A_n, the more Y is B" is studied in details in ([DGP94]).

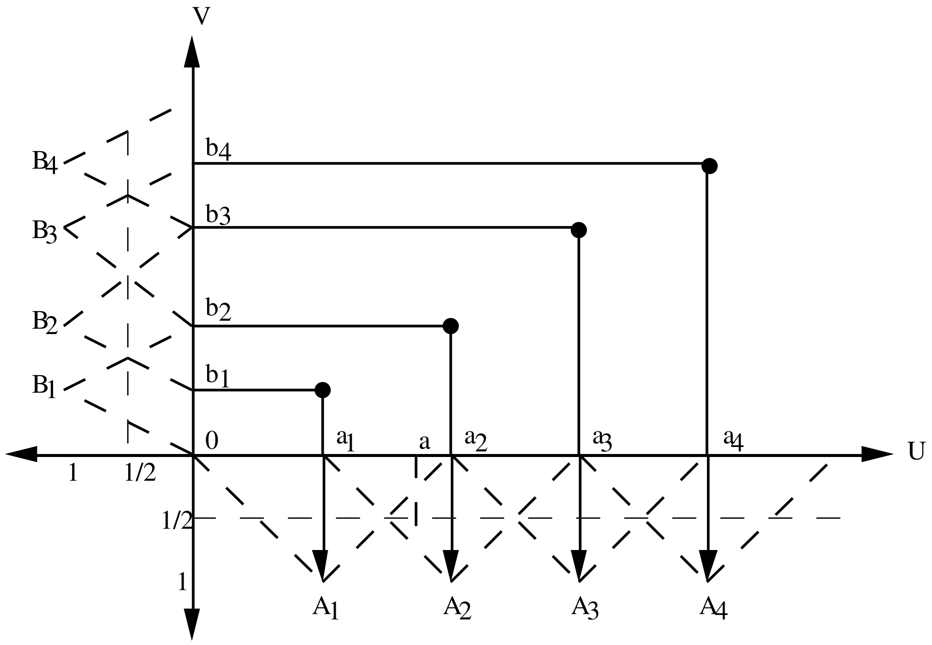


Figure 1

As we can see, these justifications of Mamdani's and Sugeno's methods, although they both rely on some closeness notion, are of different nature. The second one views interpolation as a conjunctive compromise between closeness constraints with respect to known points of the relation linking X and Y, and leads to a precise result. By contrast, Mamdani's approach is recovered through extensionality assumptions expressing that if two values are related, each one is also somewhat related to the neighbors of the other, which leads to an output which is a closeness-weighted disjunctive aggregation of the fuzzy sets appearing in the conclusion parts of the rules; the result is then a fuzzy range of possible values to be defuzzified, and interpolation is carried out only at the defuzzification level.

2.3 Qualitative Reasoning

Let us now turn towards another type of reasoning where proximity relations are explicitly handled. Let us for instance consider the two pieces of information "X is approximately equal to Y" and "Y is much larger than Z". What can be said about Z with respect to X? This kind of situation can be for instance encountered in temporal reasoning when we know that two dates are close, while one of them takes place much after a third one. An approximate equality between u and v can be modelled by a fuzzy relation E of the form $\mu_E(u,v) = \mu_L(u - v)$ which only depends on the values of the difference u – v, for instance

$$\forall u, \forall v, \mu_E(u,v) = \max\left(0, \min\left(1, \frac{\delta + \varepsilon - |u - v|}{\varepsilon}\right)\right) = \begin{cases} 1 \text{ if } |u - v| \leq \delta \\ 0 \text{ if } |u - v| > \delta + \varepsilon \\ (\delta + \varepsilon - |u - v|) / \varepsilon \text{ otherwise} \end{cases}$$

where δ and ε are respectively positive and strictly positive parameters which modulate the approximate equality. Following the same line, a more or less strong inequality can be modelled for instance by a relation I of the form

$$\forall v, \forall w, \mu_I(v,w) = \mu_K(v - w) = \max\left(0, \min\left(1, \frac{v - w - \lambda}{\rho}\right)\right) = \begin{cases} 1 \text{ if } v > w + \lambda + \rho; \lambda \geq 0 \\ 0 \text{ if } v < w + \lambda \quad ; \rho > 0 \\ (v - w - \lambda) / \rho \text{ otherwise.} \end{cases}$$

The values of λ and ρ act on the meaning of the inequality and can express shades such as 'slightly greater', 'much greater', etc. When $\lambda = 0$ and $\rho \to 0$ the usual inequality relations '>' or '≥' are recovered (depending on how $\lim_{\rho \to 0} (v - w) / \rho$ is defined for v – w goes to 0). The composition of E and I yields

$$\begin{aligned} \forall u, \forall w, \mu_{E \circ I}(u,w) &= \sup_{v \in \mathbb{R}} \min(\mu_L(u - v), \mu_K(v - w)) \\ &= \sup_{\substack{s,t \\ u-w=s+t}} \min(\mu_L(s), \mu_K(t)) \\ &= \mu_{L \oplus K}(u - w) \end{aligned}$$

where we recognize the expression of the extended sum of K and L, applying (6) to f(s,t) = s + t. Thus the possible values of the difference x – z is restricted by $L \oplus K$. This result is represented in Figure 2 where the relations defined above are used. We see that it is certain that $x \geq z + \lambda - (\delta + \varepsilon)$, since the support of $L \oplus K$ is the interval $(\lambda - \delta - \varepsilon, +\infty)$ and that the value of the difference $x - z$ belongs to $L \oplus K$ at the degree 1 as soon as $x \geq z + \lambda + \rho - \delta$. Then depending on the respective values of the parameters, x is still greater than z (but maybe not as such as y with respect to z) (if $\lambda - \delta - \varepsilon > 0$), or we are only sure that x is not much smaller than z (if $\lambda + \rho - \delta > 0$).

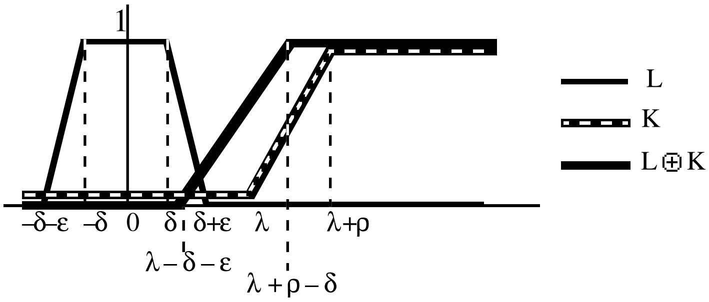


Figure 2

As it is suggested by this example, it is possible and even simple to derive inference rules in the case of fuzzy relations defined in terms of difference of values in order to perform these types of computation at the symbolic level, taking advantage of the fact that the composition of fuzzy relations then reduces to a simple arithmetic computation on the fuzzy numbers which express the semantics of the relations. This can be done as well if we use the quotient rather than the difference for comparing numbers as we are going to see.

A good example of reasoning with proximity relations is provided by qualitative reasoning on orders of magnitude, where the ideas of approximate equalities and negligibility play an important role. This can be captured in the following way ([DP89]). The relationship "X is close to Y" (between real values) is understood as "the quotient X/Y is close to 1", in the sense that the possible values of X/Y are restricted by a fuzzy set P such that $\mu_P(1) = 1$, modelling "around 1". Similarly "Z is negligible with respect to Y" will be modelled by 1 + Z/Y is close to 1, where 'close' refers here to a fuzzy set P', possibly different from P, but still such that $\mu_{P'}(1) = 1$. In other words, we define a proximity relation S(P) such that

$$\mu_{S(P)}(u,v) = \mu_P(u/v) \tag{10}$$

where P is a fuzzy number which satisfies $\mu_P(1) = 1$, $\mu_P(t) = \mu_P(1/t)$ (to ensure the symmetry of S(P)). Negligibility is expressed by

$$\mu_{N(P)}(u,v) = \mu_P(1 + u/v). \tag{11}$$

Then, given two relations S(P) and S(P'), or N(P) and N(P'), or S(P) and N(P') their composition results in a simple fuzzy arithmetic calculus on P and P'. Indeed

$$\begin{aligned}\mu_{S(P)\circ S(P')}(u,w) &= \sup_v \min(\mu_P(u/v), \mu_{P'}(v/w)) \\ &= \sup_{(s,t):s\cdot t=u/w} (\mu_P(s), \mu_{P'}(t)) = \mu_{P\odot P'}(u/w)\end{aligned}$$

where $\odot$ denotes the extension of product to fuzzy numbers. Thus,

$$S(P) \circ S(P') = S(P \odot P'). \tag{12}$$

This expresses that x is close to y in the sense of $S(P \odot P')$. It accounts for the degradation of the transitivity since $P \odot P \supset P$ usually and thus $S(P \odot P) \supset S(P)$. Similarly, it can be established that

$$N(P) \circ N(P') = N((P \ominus 1) \odot (P' \ominus 1) \oplus 1) \tag{13}$$

$$S(P) \circ N(P') = N(P \odot (P' \ominus 1) \oplus 1) \tag{14}$$

where $\ominus$ and $\oplus$ denote extended subtraction and addition. It can be shown that there is, as expected, a strengthening of the negligibility relation obtained in (13) (i.e., for $P' = P$, $N(P) \supseteq N((P \ominus 1) \odot (P \ominus 1) \oplus 1)$ and a weakening obtained in (14) (i.e., $N(P) \subseteq N(P \odot (P \ominus 1) \oplus 1))$. It is thus possible to develop inference rules where the fuzzy sets which parametrize the closeness and negligibility relations are manipulated at a symbolic level. See ([DP89], [DP91a]) for details, and fuzzy set-based counterparts of approaches proposed in the qualitative reasoning literature.

3 UNCERTAINTY MANAGEMENT

Zadeh [Zad78a] has been the first to propose the idea of representing an incomplete state of knowledge by means of a fuzzy set. If for instance we only know about quantity X that 'X is large', it means that the possible values for X are those compatible with the meaning of 'large' in the considered context, represented by the membership function of a fuzzy set, i.e., $\pi_X = \mu_{large}$ where π_X denotes the possibility distribution describing the more or less possible values for X according to what is known. As it is recalled in the next section, two measures of uncertainty, called possibility and necessity are associated with a possibility distribution. These measures turn out to be a convenient tool for the modeling of uncertainty which allows for the representation of imprecise pieces of information or partial states of ignorance. The lack of certainty in 'not A' does not entail the certainty of A in possibility theory, as it is the case with probability where Prob(not A) = 0 $\Rightarrow$ Prob(A) = 1.

3.1 Background

Let U be a set that represents the range of a variable X. Usually X stands for *the* unknown value taken by some single-valued attribute applied to an object under consideration. For instance X refers to the age of a man named Peter. A possibility distribution π_X on U is a mapping from U to the unit interval [0,1] attached to the single-valued variable X. The function π_X represents a flexible restriction which constrains the possible values of X according to the available information, with the following conventions:

$\pi_X(u) = 0$ means that X = u is definitely impossible,
$\pi_X(u) = 1$ means that absolutely nothing prevents that X = u

Note that $\pi_X(u) = 1$ and $\pi_X(u') = 1$, with $u' \neq u$, are allowed at the same time. Intermediary levels of plausibility about the possible values of X are modelled by letting $\pi_X(u)$ between 0 and 1 for some values u. The quantity $\pi_X(u)$ thus represents the degree of possibility of the assignment X = u. Thus we can acknowledge the fact that some values u are more possible than others, according to what is known. In other words, if we consider the logical propositions p_u whose semantic contents is 'X = u', π_X encodes a *preference* relation among these propositions describing possible states of the world: if $\pi_X(u) < \pi_X(u')$, u' is strictly preferred to u as a possible candidate for the value of X according to the available information, among more or less possible (or allowed) values, i.e., among the u's such that $\pi_X(u) > 0$. Clearly, if U exhaustively contains all the values which may be thought of for X, at least one of the elements of U should be fully possible as a value of X, so that $\exists u, \pi_X(u) = 1$ (*normalization*). To summarize, a possibility distribution π_X encodes a preference relation about the possible values of the variable X, acording to the available knowledge.

Given a possibility distribution π_X, several set-valued functions of interest can be defined in order to estimate the tendency of X to belong to a given subset A of U. Two of these set-functions are well-known and widely used, namely

– the so-called measure of possibility ([Zad78a])

$$\Pi(A) = \sup_{u \in A} \pi_X(u) \tag{15}$$

which estimates the *consistency* of the statement 'X ∈ A' with what we know about the possible values of X; hence we look for the element(s) of A having the greatest possibility degree according to π_X (since A is all the more possible as there exists in A a value with a high degree of possibility)

– the dual measure of necessity [Zad79b][DP80]

$$\begin{aligned} N(A) &= 1 - \Pi(\bar{A}) && (16)\\ &= \inf_{u \notin A} 1 - \pi_X(u) && (17) \end{aligned}$$

which estimates to what extent each value outside A, i.e., in the complement $\bar{A}$ of A, has a low degree of possibility, thus the values with the higher degrees of possibility should be among the elements of A, which makes us somewhat *certain* that indeed X belongs to A. The duality relation (16) between Π and N expresses that A is all the more certain as $\bar{A}$ is less consistent with the available knowledge. It assumes that π_X is normalized. The normalization of π_X entails $\max(\Pi(A),\Pi(\bar{A})) = 1$, and then $N(A) > 0$ entails $\Pi(A) = 1$, i.e., A should be completely consistent with what is known before being somewhat certain.

Note that, in agreement with the idea that only the ordering among the possibility degrees (and then among the possible values) is meaningful, the set functions Π and N only use the qualitative operations 'sup' and 'inf' (or respectively 'max' and 'min' when U is finite) and the order reversing operation $1 - (\cdot)$.

Apart from Π and N, two other quantities can be defined with the same operations, namely [DP92a]

– a measure of 'guaranteed possibility'

$$\Delta(A) = \inf_{u \in A} \pi_X(u) \tag{18}$$

which estimates to what extent *all* the values in A are actually possible for X according to what is known, i.e. any value in A is at least possible for X at the degree $\Delta(A)$. Clearly Δ is a stronger measure than Π, i.e., $\Delta \leq \Pi$, since Π only estimates the existence of at least one value in A compatible with the available knowledge, while the evaluation provided by Δ concerns *all* the values in A. Note also that Δ and N are unrelated, since N(A) estimates the certainty that the value of X is in A by checking the impossibility of all the values out of A, while $\Delta(A) > 0$ considers the possibility of all the values in A.

– a dual measure of 'potential certainty'

$$\nabla(A) = 1 - \Delta(\bar{A}) \tag{19}$$

$$= \sup_{u \notin A} 1 - \pi_X(u) \tag{20}$$

which estimates to what extent there exists at least one value in the complement of A which has a low degree of possibility; this is a necessary condition for having '$X \in A$' somewhat certain (but in general far from being sufficient, except if $\bar{A}$ has only one element). Obviously we have $N \leq \nabla$. The duality relation (19) expresses that the potential certainty of A corresponds to the absence of guaranteed possibility for $\bar{A}$. If we further have the information that there exists at least one value in U which is impossible for X (in other words, there is not a complete uncertainty about the value of X), i.e. $\exists\, u \in U$, $\pi_X(u) = 0$, or equivalently $1 - \pi_X(u) = 1$, we have the constraint $\min(\Delta(A),\Delta(\bar{A})) = 0$ and then $\Delta(A) > 0$ entails $\nabla(A) = 1$. At the technical level, it is always possible to add an element to U if necessary in order to have $1 - \pi_X$ normalized. Δ and ∇ are monotonically decreasing set functions (in the wide sense) with respect to set inclusion; it contrasts with Π and N which are monotonically increasing.

Note that the four quantities $\Pi(A)$, N(A), $\Delta(A)$ and $\nabla(A)$ are only weakly related since they are only constrained by

$$\max(\Pi(A), 1 - N(A)) = \sup_{u \in U} \pi_X(u) \qquad (= 1 \text{ if } \pi_X \text{ is normalized})$$
$$\min(\triangle(A), 1 - \triangledown(A)) = \inf_{u \in U} \pi_X(u) \qquad (= 0 \text{ if } 1 - \pi_X \text{ is normalized})$$

together with the duality relations (16) and (19). As it can be seen, the four quantities $\Pi(A)$, $\Delta(A)$, $N(A)$, $\nabla(A)$ summarize all the information which can be extracted from π_X by taking the supremum or the infimum of π_X over A or over its complement.

A possibility distribution is not always specified as such, but often through the qualification of subsets of U. Let A be an *ordinary* subset of U. Two kinds of specification exist, namely

i) we know that it is (completely) *certain* that the value of X lies in A (for short "A is certain for X"). It means that any value outside A is (completely) impossible, i.e. $\forall u \notin A$, $\pi_X(u) = 0$ and π_X is unspecified over A, or if we prefer

$$\text{"A is certain for X" is translated by } \forall u \in U, \pi_X(u) \leq \mu_A(u) \qquad (21)$$

where μ_A is the $\{0,1\}$-valued characteristic function of A;

ii) A is a (completely) possible range for X (for short "A is possible for X"). It means that $\forall u \in A$, $\pi_X(u) = 1$ and π_X remains unspecified outside of A, or equivalently that

$$\text{"A is possible for X" is translated by } \forall u \in U, \mu_A(u) \leq \pi_X(u). \qquad (22)$$

In the first kind of specification we express our certainty, delimiting a subset A which for sure contains the value of X; the larger A, the more imprecise we remain about the value of X, the smaller A, the better the information. On the contrary, in the second kind of specification, we rather express our uncertainty, stating a range A of values which are indeed possible for X without rejecting any value outside of A; in some sense we are recognizing the extent of our ignorance with respect to the precise value of X. Then, the smaller A, the less informative the piece of knowledge; the larger A, the more we know *for sure* the imprecision of our state of knowledge.

Let us now consider the cases of qualification where the possibility or the certainty is not complete but corresponds to an intermediary level α in the scale [0,1]. It leads to the two following generalizations of (21) and (22)

i) The statement "it is certain at least at the degree α that the value of X is in A", will be interpreted as "any value outside A is at most possible at the complementary degree, namely $1 - \alpha$", i.e. $\forall\ u \notin A, \pi_X(u) \leq 1 - \alpha$, which leads to

"A is α-certain for X" is translated by $\forall\ u \in U, \pi_X(u) \leq \max(\mu_A(u), 1 - \alpha)$ (23)

It can be checked that this is equivalent to $N(A) \geq \alpha$. Note that for $\alpha = 1$, (21) is recovered. When α decreases from 1 to 0, our knowledge evolves from complete certainty in A to acknowledged ignorance about X.

ii) The statement "A is a possible range for X at least at the degree α" will be understood as $\forall\ u \in A, \pi_X(u) \geq \alpha$, which leads to

"A is α-possible for X" is translated by $\forall\ u \in U, \min(\mu_A(u), \alpha) \leq \pi_X(u)$. (24)

It can be checked that this is equivalent to $\triangle(A) \geq \alpha$. Note that for $\alpha = 1$, (22) is recovered. When α decreases from 1 to 0, our knowledge evolves from the certainty that A is the minimal range of our ignorance to a total lack of information whatsoever. This kind of possibility-qualification goes back to [Zad78b]) and [San78].

The above representations (23) and (24) can be generalized to the case where A is a *fuzzy* set, namely we have [DP92b]

i) $\forall\ u \in U, \pi_X(u) \leq \max(\mu_A(u), 1 - \alpha)$ (α-certainty qualification)
ii) $\forall\ u \in U, \pi_X(u) \geq \min(\mu_A(u), \alpha)$ (α-possibility qualification)

It is still equivalent to $N(A) \geq \alpha$ and $\triangle(A) \geq \alpha$ provided we use the following appropriate extensions of N and $\triangle$ when A is fuzzy

$$N(A) = \inf_{u \in U} (1 - \mu_A(u)) \rightarrow (1 - \pi_X(u)) \quad (25)$$

$$\triangle(A) = \inf_{u \in U} \mu_A(u) \rightarrow \pi_X(u) \quad (26)$$

where $\rightarrow$ is the so-called Gödel's implication $r = s \rightarrow t = \begin{cases} 1 \text{ if } x \le t \\ t \quad s > t \end{cases}$. Indeed we have the equivalences $\max(s, 1 - r) \ge t \Leftrightarrow r \le (1 - s) \rightarrow (1 - t)$ and $\min(s,r) \le t \Leftrightarrow r \le s \rightarrow t$, which are the basis of fuzzy relational equation solving ([DSPS89]). When A is a crisp subset (25) and (26) reduce to (17) and (18) respectively.

3.2 Uncertain Fuzzy Rules

Certainty- and possibility-qualification provide a tool for distinguishing between different intended meanings that a fuzzy rule of the form "if X is A then Y is B" may convey. Indeed such a rule can be interpreted in various ways according to how the if-part qualifies the then-part. Since the satisfaction of the condition part by a precise state of fact $X = x_0$ is a matter of degree when A is fuzzy, this degree may act as a lower bound of the certainty or the possibility of the conclusion part of the rule. Thus, the expressions of the α-certainty and the α-possibility qualification give birth to two kinds of fuzzy rules when the coefficient α becomes dependent on the evaluation of some condition. First we consider rules which relate a variable X ranging on U to a variable Y ranging on V, and which express that "the more X is A, the more certain Y lies in B". It may mean something like "$\forall$u, if X = u, it is at least $\mu_A(u)$-certain that B is a range for Y". This can be represented by the following constraint on the conditional possibility distribution $\pi_{Y|X}$ relating the possible values of X and Y, due to (23)

$$\forall\, u \in U, \forall\, v \in V, \pi_{Y|X}(v,u) \le \max(\mu_B(v), 1 - \mu_A(u)) \tag{27}$$

Such a rule will be called "*certainty rule*". A second kind of fuzzy rule called "*possibility rule*" expresses that "the more X is A, the more possible B is a range for Y" and which can be understood as "$\forall$u, if X = u, it is at least $\mu_A(u)$-possible that Y lies in B". It corresponds to the following constraint on the conditional possibility distribution $\pi_{Y|X}$ representing the rule

$$\forall\, u \in U, \forall\, v \in V, \min(\mu_A(u), \mu_B(v)) \le \pi_{Y|X}(v,u) \tag{28}$$

due to (24). This model of fuzzy rule is close to Mamdani [Mam77]'s original proposal in fuzzy logic control. Note that in the particular case where A is an ordinary subset and where we know that, if X is in A, B is both a possible and a certain range for Y, (27) and (28) yield

$$\begin{cases} \forall\ u \in\ A,\ \pi_{Y|X}(v,u) = \mu_B(v) \\ \forall\ u \notin\ A,\ \pi_{Y|X}(v,u) \text{ is completely unspecified.} \end{cases} \tag{29}$$

This corresponds to the usual modelling of a fuzzy rule with a non-fuzzy condition part.

Possibility rules also read "the more X is A, the more it is possible that Y takes values v which have high degree of membership in B". Reversing the roles of the expressions "high degree of membership" and "possible values" in this rule leads to the new rule "the more X is A, the higher the degree of membership in B of the possible values of Y", or if we prefer, more briefly "the more X is A, the more Y is B". The constraint which represents this new kind of rule, called "*gradual rule*", is thus obtained in exchanging μ_B and $\pi_{Y|X}$ in (28). It yields the inequality

$$\forall\ u \in\ U,\ \forall\ v \in\ V,\ \min(\mu_A(u),\ \pi_{Y|X}(v,u)) \leq \mu_B(v) \tag{30}$$

When $\pi_{Y|X}(v,u)$ takes only the values 0 or 1, we recognize the particular case of gradual rules already encountered in Section 2.2. More precisely, the intended meaning of a gradual rule can be easily read in (30): "the greater the degree of membership of the value of X to the fuzzy set A and the more the value of Y is possibly related to X, the greater should be the degree of membership to B for this value of Y". Using again the equivalence $\min(a,t) \leq b \Leftrightarrow t \leq a \rightarrow b$, (30) can be rewritten under the form

$$\forall\ u \in\ U,\ \forall\ v \in\ V,\ \pi_{Y|X}(v,u) \leq \mu_A(u) \rightarrow \mu_B(v) = \begin{cases} 1 & \text{if } \mu_A(u) \leq \mu_B(v) \\ \mu_B(v) & \text{if } \mu_A(u) > \mu_B(v) \end{cases} \tag{31}$$

Fuzzy rules based on (30) (with possibly combination different from the minimum) have been especially considered in the past by Trillas and Valverde ([TV85]), Bouchon ([Bou87]) and the authors.

Exchanging similarly μ_B and $\pi_{Y|X}$ in (27), we get another kind of gradual rule represented by

$$\forall\, u \in U, \forall\, v \in V, \max(\pi_{Y|X}(v,u), 1 - \mu_A(u)) \geq \mu_B(v) \tag{32}$$

$$\Leftrightarrow 1 - \mu_B(v) \geq \min(\mu_A(u), 1 - \pi_{Y|X}(v,u))$$

$$\Leftrightarrow \pi_{Y|X}(v,u) \geq \begin{cases} 0 & \text{if } \mu_A(u) + \mu_B(v) \leq 1 \\ \mu_B(v) & \text{if } \mu_A(u) + \mu_B(v) > 1 \end{cases}$$

using the equivalence $\min(a, 1 - t) \leq 1 - b \Leftrightarrow t \geq 1 - (a \to (1 - b)) = a \curlywedge b$ where $\curlywedge$ denotes a non-commutative conjunction. As it can be seen above on the second expression of (32), this gradual rule expresses that "the more X is A and the less Y is related to X, the less Y is B".

The particular case of (30) where $\pi_{Y|X}$ reduces to a function f such that $\pi_{Y|X}(v,u) = 1$ if $v = f(u)$ and $\pi_{Y|X}(v,u) = 0$ otherwise, i.e. $\forall\, u \in U, \mu_A(u) \leq \mu_B(f(u))$, has been considered for a long time in the fuzzy set literature under the name of "fuzzy function", see [NR75]). More generally, viewing $\pi_{Y|X}$ as encoding a fuzzy relation R, i.e. $\pi_{Y|X}(v,u) = \mu_R(u,v)$, and noticing that (30) is equivalent to $\forall\, v \in V, \sup_u \min(\mu_A(u), \mu_R(u,v)) \leq \mu_B(v)$, we recognize that it expresses the fuzzy set inclusion $A \circ R \subseteq B$, where $A \circ R$ is the image of the fuzzy set A by the fuzzy relation R. When A, B, R are *non-fuzzy*, we have the equivalence $A \circ R \subseteq B \Leftrightarrow R \subseteq \bar{A} + B$ (+ denotes the Cartesian co-product $A + B = \overline{\bar{A} \times \bar{B}}$) ; this equivalence no longer holds in the fuzzy case and its right part is the counterpart of (27). Similarly, in the non-fuzzy case we have the equivalence $A \circ \bar{R} \subseteq \bar{B} \Leftrightarrow R \supseteq A \times B$ which no longer holds in the fuzzy case ; the left and right parts respectively correspond to (32) and (28) and clearly (32) expresses that any element of U which is not in relation R with an element of A should be outside B.

These four types of rules correspond to modifications of μ_B by a function $\mu_{\tau(u)}$ from [0,1] to [0,1] which may interpreted as a fuzzy truth-value $\tau(u)$ ([Zad78b]), namely

for (27) and (30)-(31): $\pi_{Y|X}(v,u) \leq \mu_{\tau(u)}(\mu_B(v))$;
for (28) and (32): $\pi_{Y|X}(v,u) \geq \mu_{\tau(u)}(\mu_B(v))$

where $\mu_{\tau(u)}$ in each case depends on $\mu_A(u)$. This possibility of interpreting fuzzy inference by means of fuzzy truth-values acting as modifiers has been noticed for a long time ([Tsu79] ; [Bal79]). The corresponding modifier functions $\mu_{\tau(u)}$ are given for the four types of rules in Figure 3 with $\alpha = \mu_A(u)$.

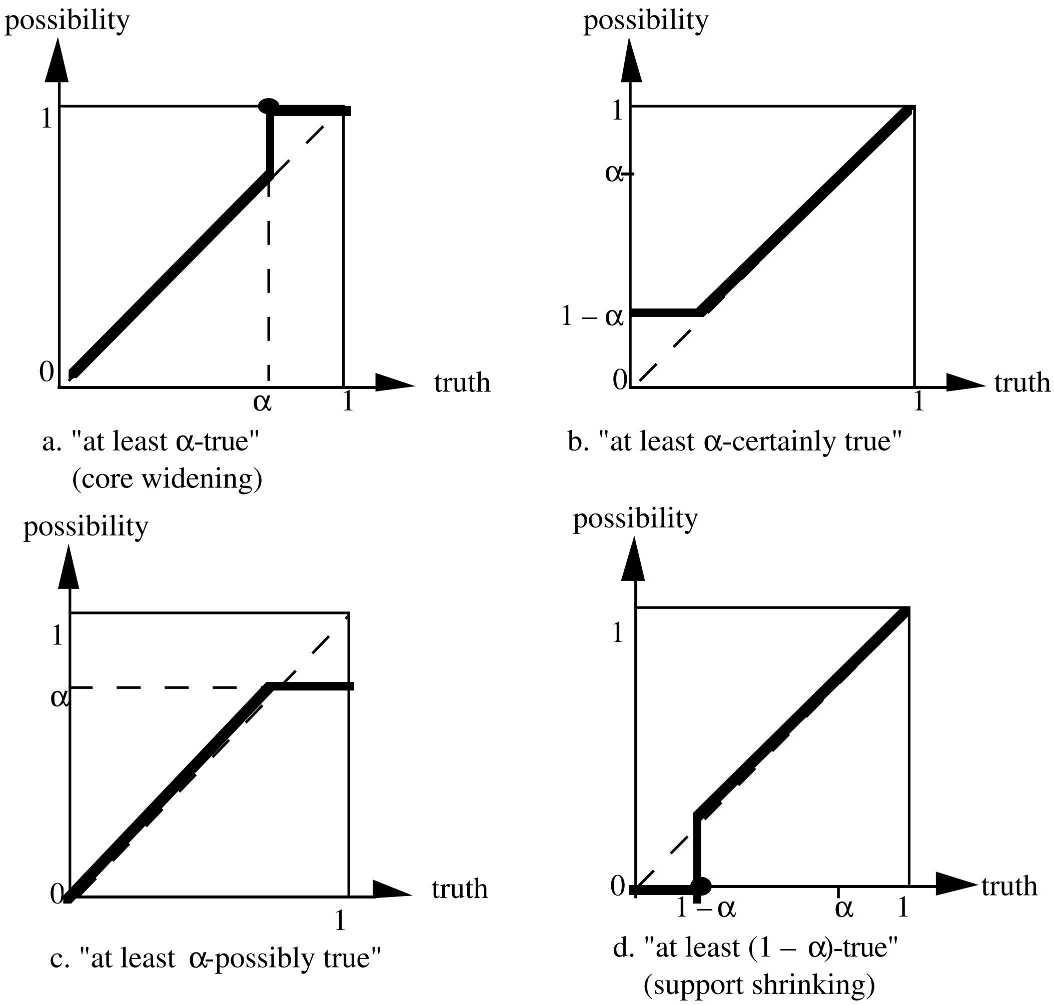


Figure 3. Four basic modifier functions

As it appears in Figure 3.a, a gradual rule of the type (30)-(31) can be interpreted in the following way : $\forall$ u $\in$ U, if X = u then Y is B is at least $\mu_A(u)$-true. It is not a crisp "at least α-true" (which would correspond to the ordinary subset $[\alpha,1]$), but a fuzzy one in agreement with the fact that $\mu_\tau(t) = t$, $\forall$ t $\in$ [0,1] is the identity modifier function accounting for the truth value "true". To summarize, in case of

– *possibility rules*, the possibility distribution $\pi_{Y|X}(\cdot,u)$ is bounded from below by $\varphi(\mu_B)$ with $\varphi(t) = \min(\mu_A(u), t)$, i.e. B is truncated up to the height $\alpha = \mu_A(u)$;
– *certainty rules*, the possibility distribution $\pi_{Y|X}(\cdot,u)$ is bounded from above by $\varphi(\mu_B)$ with $\varphi(t) = \max(t, 1 - \mu_A(u))$, i.e. B is drowned in a level of indetermination $1 - \alpha$;

– *gradual rules* (core widening), the possibility distribution $\pi_{Y|X}(\cdot,u)$ is bounded from above by $\varphi(\mu_B)$ with $\varphi(t) = \mu_A(u) \to t$ (where $\to$ denotes Gödel's implication), i.e. the core of B is enlarged ;
– *gradual rules* (support shrinking), the possibility distribution $\pi_{Y|X}(\cdot,u)$ is bounded from below by $\varphi(\mu_B)$ with $\varphi(t) = \mu_A(u) \wedge t = 0$ if $\mu_A(u) + t \leq 1$ and $\varphi(t) = t$ otherwise, i.e. the support of B is reduced, truncated.

Statements of rules involving "the less" in place of "the more" are easily obtained by changing A or B into their complements $\bar{A}$ and $\bar{B}$, due to the equivalence between "the more X is A" and "the less X is $\bar{A}$" (with $\mu_{\bar{A}} = 1 - \mu_A$).

3.2 Approximate Reasoning with Fuzzy Rules

3.2.1 Information Principles

In the previous section we have seen how the representation of fuzzy rules and more generally certainty and possibility qualifications lead to inequality constraints on possibility distributions. In case of certainty qualification, the inequality is of the type $\pi_X \leq \mu_A$ where μ_A may be any [0,1]-valued (normalized) membership function ; it includes the case where μ_A is of the form $\mu_A = \max(\mu_F, 1 - \alpha)$, i.e. "F is α-certain for X" is equivalent to "A is certain for X" (where A is necessarily a fuzzy set). The *principle of minimal specificity* stipulates that each value u in the domain U of a variable X (in the general case X is a vector and U a Cartesian product) should receive the largest degree of possibility which is in agreement with the constraint(s) ; i.e. it leads to take $\forall\, u \in U$, $\pi_X(u) = \mu_A(u)$ in the above case. Indeed choosing a particular π such that $\pi < \mu_A$ in order to represent our knowledge about X would be arbitrarily too precise. The principle of minimal specificity has a role in possibility theory similar to the use of the maximal entropy principle in probability theory. The notion of specificity of a possibility distribution has been introduced by Yager ([Yag83]). The absence of constraint corresponds to the situation of complete ignorance modelled by $\forall\, u \in U$, $\pi_X(u) = 1$. At the opposite, complete knowledge corresponds to a maximal constraint, such that $\exists u_0$, $A = \{u_0\}$, i.e. $\pi_X(u_0) = 1$ and $\pi_X(u) = 0$, $\forall\, u \neq u_0$. In that sense any possibility distribution π_X is provisional in nature and likely to be improved by further information, when the available one is not complete. When several pieces of information constraining the possible values of X become available, i.e. $\pi_X \leq \mu_{A_i}$, i = 1,n, the principle of minimum specificity leads to assume

$$\pi_X = \min_{i=1,\ldots,n} \mu_{A_i} \tag{33}$$

as a straightforward consequence of the inequalities. Note that the complete reliability of the sources entails that $\sup_u \min_{i=1,\ldots,n} \mu_{A_i}(u) = 1$. Otherwise, the sources are conflicting as soon as $\pi_X(u) < 1$, $\forall u$. Hence subnormalization corresponds to *partial inconsistency*. Clearly if we have two constraints $\pi_X \leq \mu_{A_1}$ and $\pi_X \leq \mu_{A_2}$, such that $\mu_{A_1} \leq \mu_{A_2}$, (i.e., we have the fuzzy set inclusion $A_1 \subseteq A_2$) the second piece of information is redundant and A_1 is said to be more specific than A_2.

Obviously a different information principle needs to be applied in case of possibility qualification where we deal with inequalities of the type $\pi_X \geq \mu_A$. Here μ_A does not need to be normalized, indeed μ_A may be of the form $\mu_A = \min(\alpha, \mu_F)$, but as already said it is expected that $\exists\, u \in U$, $\mu_A(u) = 0$. A *principle of maximal specificity* can be advocated in this case in order to limit the scope of the possible values for X to the values which are indeed known as (somewhat) possible. While the principle of minimal specificity embodies the commonsense claim that anything that is not impossible should not be ruled out, the converse principle expresses that anything which is not established as being possible can be neglected. The role of information principles in possibility and other non-standard theories of uncertainty is particularly stressed in the book by Klir and Folger ([KF88]).

3.2.2 Combination and Projection

Let us now assume that there are two variables X and Y ranging on U and V respectively, with possibility distributions π_X and π_Y respectively. The principle of minimal specificity leads us to define the joint possibility distribution $\pi_{X,Y}$ as the *combination*

$$\pi_{X,Y} = \min(\pi_X, \pi_Y) \tag{34}$$

as a consequence of the inequalities $\pi_{X,Y}(u,v) \leq \pi_X(u)$, $\forall$ v and $\pi_{X,Y}(u,v) \leq \pi_Y(v)$, $\forall$u, which express that the possibility that X = u *and* Y = v is necessarily upper-bounded by the possibility that X = u (nothing being said about Y) on the one hand

and the possibility that Y = v on the other hand. Note that (34) is in agreement with (33). Indeed the possibility distribution π_X on U (resp. π_Y on V) can be extended to U $\times$ V into a possibility distribution $\pi^1{}_{X,Y}$ (resp. $\pi^2{}_{X,Y}$) by stating that $\forall$u, $\forall$v, $\pi^1{}_{X,Y}(u,v) = \pi_X(u)$ (resp. $\pi^2{}_{X,Y}(u,v) = \pi_Y(v)$), since nothing is said about Y (resp. X) ; this is called *cylindrical extension*. Then $\pi_{X,Y} = \min(\pi^1{}_{X,Y},\pi^2{}_{X,Y})$ from (33). As it can be seen it is equivalent to first apply the minimal specificity principle for each possibility distribution representing a piece of information and then to combine by (34), or to first combine the constraints corresponding to certainty-qualified statements and then to apply the minimal specificity principle to the result.

Conversely, given a possibility distribution $\pi_{X,Y}$ restricting the possible values of a pair of variables (X,Y) the induced restriction on the possible values of X (resp. Y) can be obtained by the *projection* of $\pi_{X,Y}$ on U (resp. V), i.e.

$$\pi_X(u) = \sup_{v\in V} \pi_{X,Y}(u,v) \ \ ; \ \ \pi_Y(v) = \sup_{u\in U} \pi_{X,Y}(u,v) \qquad (35)$$

Again this is in agreement with the principle of minimal specificity since we compute the largest possibility degrees satisfying the constraint X = u, $\forall$ u $\in$ U (resp. Y = v, $\forall$ v $\in$ V). The combination can be straightforwardly extended to n-tuples of variables and the projection can be performed on any subset of such tuples. Combination and projection as defined above are the basis of the theory of approximate reasoning introduced by Zadeh ([Zad79a]). The combination and projection operations in possibility theory play roles which are respectively similar to the construction of a joint distribution (under a stochastic independence assumption) and to the computation of a marginal distribution in probability theory. The simultaneous satisfaction of (34) and (35) supposes that X and Y are logically independent (i.e. that the degree of possibility that X = u does not depend on the value of Y, $\forall$u, and similarly exchanging X and Y).

Note also that viewing fuzzy relations as possibility distributions, the composition of fuzzy relations (see Section 2.1) can be interpreted as a combination followed by a projection in the above sense.

3.2.3 Generalized Modus Ponens

An important particular case of the application of the combination/projection principle is the generalized modus ponens where a factual piece of information X is A' is combined with a fuzzy rule relating X and Y, in order to get a conclusion of the form Y is B', after projection. We only consider the cases of certainty and possibility rules in the following. Inference with gradual rules has been already presented in Section 2.2 in the particular case where A' = $\{x_0\}$ corresponds to a precise input. See [DMP88] for computational issues in the cases of a fuzzy input A' and of rules with compound conditions.

- <u>certainty rules</u>: From the two pieces of information "X is A' " and "the more X is A, the more certain Y is B", i.e. $\pi_X = \mu_{A'}$, $\pi_{Y|X}(v,u) = \max(\mu_B(v), 1 - \mu_A(u))$, applying (27), we obtain by combination and projection

$$\begin{aligned} \pi_Y(v) &= \sup_u \min(\mu_{A'}(u), \max(\mu_B(v), 1 - \mu_A(u)) \\ &= \max(\mu_B(v), 1 - N_{A'}(A)) \qquad (36) \end{aligned}$$

  where $N_{A'}(A)$ is a necessity measure of the fuzzy event A, defined from the possibility distribution $\pi = \mu_{A'}$ (not to be confused with (25)), such that $N_{A'}(A) = 1$ if and only if the support of A' is included in the core of A, i.e. $\{u, \mu_{A'}(u) > 0\} \subseteq \{u, \mu_A(u) = 1\}$. Hence (36) expresses that the conclusion "Y is B" is all the more certain as the fuzzy set A' of possible values of X corresponds to elements highly compatible with A. The use of certainty rules thus lead to the propagation of uncertainty coefficients.

- <u>possibility rules</u>: Let us consider the generic situation of two objects described in terms of attribute variables X and Y. The indices 1 and 2 will refer to the values of these attributes for each of these objects respectively. Let us suppose we have the following information : $\pi_{X_1} = \mu_{A_1}$, $\pi_{Y_1} = \mu_{B_1}$, $\pi_{X_2} = \mu_{A_2}$ and the rule "the more similar the values of X_1 and X_2, the more possible the approximate equality of the values of Y_1 and Y_2" (e.g., the more similar two second-hand cars, the more possible they have approximately equal prices). Let S and T be the fuzzy relations "similar" and "approximately equal" in the above rule. Then by combination/ projection we obtain for Y_2, applying (28)

$$
\begin{aligned}
\pi_{Y_2}(v_2) &= \sup_{u_1,u_2,v_1} \min(\mu_{A_1}(u_1), \mu_{A_2}(u_2), \mu_{B_1}(v_1), \mu_S(u_1,u_2), \mu_T(v_1,v_2)) \\
&= \min(\mu_{B \circ T}(v_1), \sup_{u_1,u_2} \min(\mu_S(u_1,u_2), \mu_{A_1}(u_1), \mu_{A_2}(u_2)) \\
&= \min(\mu_{B \circ T}(v_2), \Pi_{A_1 \times A_2}(S)) \qquad (37)
\end{aligned}
$$

where $\Pi_{A_1 \times A_2}(S)$ is the possibility measure of the fuzzy event "X_1 and X_2 are similar", given that "X_1 is A_1" and "X_2 is A_2" and (34) is used to combine μ_{A_1} and μ_{A_2}. The conclusion obtained by (37) means that the values in B or approximately equal to a value in B are all the more possible for Y_2 as X_1 and X_2 are indeed possibly similar. This kind of weak conclusion is in the spirit of analogical reasoning which can lead only to tentative conclusions.

It is worth noticing that the effect of parallel certainty rules leads to increase the precision of the response, while in the case of possibility rules, the response gets wider. This is in conformity with the semantics of the rules. Namely in the case of certainty rules, given that X is totally compatible with both A_1 and A_2, we conclude that both B_1 and B_2 are certain, i.e. $X \in B_1 \cap B_2$. However with possibility rules, if X is totally compatible with both A_1 and A_2, then both B_1 and B_2 are possible responses, i.e. $B_1 \cup B_2$ is a possible range for X. This can be easily checked by combining two parallel rules (either two certainty rules or two possibility rules) in presence of a precise input.

3.4 Possibilistic Logic

In the previous sections we have dealt with fuzzy predicates and with uncertainty measures together. In the remaining sections we focus on the use of possibility and/or necessity measures for the handling of uncertainty pertaining to classical, i.e., non-fuzzy, formulas p, q, r… (which can be only true or false).

Several approaches have been proposed for dealing with uncertainty and/or vagueness in automated reasoning; see [DP91c, Part 2] for an overview. However a large part of them are based on *fuzzy* logic, which completely departs from the *possibilistic* logic we present in this section. Fuzzy logic deals with propositions involving vague predicates (or properties whose satisfaction can be a matter of degree) and manipulates truth degrees which are truth-functional with respect to each connective, whereas possibilistic logic involves certainty and possibility degrees which are not compositional for all connectives and which are attached to classical formulae, i.e. containing only non-vague propositions or predicates (in the simplest case). The lack of complete certainty about the truth of a considered formula is to be understood as a consequence of a lack of complete information.

A possibilistic logic formula is a first order logic formula with a numerical weight between 0 and 1 which is a lower bound on a possibility measure Π or on a necessity measure N. Thus this lower bound should obey the characteristic axioms governing these measures, i.e. $\forall p$, $\forall q$, $N(p \wedge q) = \min(N(p),N(q))$ and $\Pi(p \vee q) = \max(\Pi(p),\Pi(q))$ respectively for necessity and possibility measures, with the duality relation $N(p) = 1 - \Pi(\neg p)$. We have the usual limit conditions $\Pi(\perp) = N(\perp) = 0$, $\Pi(\top) = N(\top) = 1$, where $\perp$ and $\top$ stand for the contradiction and the tautology respectively. However we only have $N(p \vee q) \geq \max(N(p),N(q))$ and $\Pi(p \wedge q) \leq \min(\Pi(p),\Pi(q))$ since for $q = \neg p$ we get $N(p \vee q) = 1$ and $\Pi(p \wedge q) = 0$, while $\max(N(p),N(\neg p))$ is less than 1 when the situation regarding the truth or falsity of p is uncertain, and $\min(\Pi(p),\Pi(\neg p))$ is strictly positive if p and $\neg p$ are both somewhat possible (state of partial ignorance). The weight attached to a formula represents to what extent it is possible or it is certain that the formula holds for true given the available information. A semantics has been proposed first when only lower bounds on a necessity measure are used ([DLP92]) and then extended to the general case where lower bounds of both possibility and necessity are allowed ([LDP91]). For the sake of brevity let us only indicate the semantics attached to necessity-valued formulas (p,α) where the intended meaning of α is $N(p) \geq \alpha$. Let $M(p)$ be the (classical) set of models of p, i.e. the set of interpretations which make p true. The semantics of (p,α) is represented by the fuzzy set of models $M(p,\alpha)$ where ω denotes an interpretation

$$\mu_{M(p,\alpha)}(\omega) = 1 \text{ if } \omega \in M(p) \ ; \ \mu_{M(p,\alpha)}(\omega) = 1 - \alpha \text{ if } \omega \notin M(p). \qquad (38)$$

In other words, the interpretations compatible with (p,α) are restricted by the above possibility distribution. The ones in $M(p)$ are considered as fully possible while the ones outside are all the more possible as α is smaller, i.e., as the piece of knowledge is less certain. This is clearly in agreement with certainty qualification as discussed in Section 3.1.

In case of several pieces of knowledge (p_i,α_i), i = 1,n forming a knowledge base $\mathcal{K}$, in agreement with the minimal specificity principle, we associate the following possibility distribution, built by combining the membership functions $\mu_{M(p_i,\alpha_i)}$, namely

$$\pi_{\mathcal{K}}(\omega) = \min_{i=1,n} \mu_{M(p_i,\alpha_i)}(\omega). \tag{39}$$

The lack of certainty in p_i, estimated by $1 - \alpha_i$, is committed to the interpretations which are not models of p_i. By performing the conjunction of the $M(p_i,\alpha_i)$'s, we associate each interpretation ω with a weight equal to $\pi_{\mathcal{K}}(\omega) = \min_{i=1,n} \mu_{M(p_i,\alpha_i)}(\omega)$. Thus the weights attached to formulas in the knowledge base $\mathcal{K} = \{(p_i,\alpha_i)\}$ induce an ordering among the interpretations (according to their level of possibility $\pi(\omega)$). It is very similar to Shoham ([Sho88])'s preferential model semantics; see ([DP91b]) on this point. It can be checked that $\forall i$, $N(p_i) = 1 - \Pi(\neg p_i) \geq \alpha_i$, with $\Pi(p_i) = \sup\{\pi_{\mathcal{K}}(\omega), \omega \in M(p_i)\}$.

The following deduction rules have been proved sound and complete for refutation with respect to the above semantics

resolution rule : $\dfrac{(\neg p \vee q, \alpha) \quad (p \vee r, \beta)}{(q \vee r, \min(\alpha,\beta))}$

particularization : $\dfrac{(\forall x P(x), \alpha)}{(P(a), \alpha)}$ (as well as more general substitutions)

They are other examples of local inferences expressed at the symbolic level.

It we want to compute the certainty degree which can be attached to a formula, we add to $\mathcal{K}$ the clause(s) obtained by refuting the proposition to evaluate with a necessity degree equal to 1. Then it can be shown that any lower bound obtained on the contradiction, by resolution, is a lower bound of the necessity of the proposition to evaluate. It can be proved that what is obtained is in complete agreement with the combination and projection approach taking into account the above semantics [DP91c, Part 2].

The introduced semantics enables us to define the degree *of partial inconsistency* of a knowledge base $\mathcal{K}$ which is equal, in the case of necessity-weighted formulas, to $\mathrm{Inc}(\mathcal{K}) = 1 - \sup_\omega \min_{i=1,n} \mu_{M(p_i,\alpha_i)}(\omega)$. Then it can be shown that this degree estimates to what extent the lower bounds in the knowledge base violate the characteristic axiom of necessity measures and to what extent the fuzzy set of models of the knowledge base is empty. When Inc($\mathcal{K}$) = 0, $\mathcal{K}$ is consistent and this is equivalent to the classical consistency of the set of formulas in $\mathcal{K}$ without taking into account the weights. It has been also shown that is it possible to reason with such partially inconsistent knowledge bases, still preserving the above-mentioned soundness and completeness results. Roughly speaking, the conclusions which can be obtained with a degree of uncertainty strictly higher than Inc($\mathcal{K}$) are still meaningful since for sure only a consistent subpart of $\mathcal{K}$ is used for deducing them. Indeed in any inconsistent sub-base of $\mathcal{K}$ there is (at least) a clause with a weight less or equal to Inc($\mathcal{K}$). In case of partial inconsistency, it is has been shown that possibilistic logic behaves in a non-monotonic way ([DP91b]). More specifically, the possibility distribution $\pi_{\mathcal{K}}$ encodes a preference relation over interpretations, and inference in possibilistic logic corresponds to a preferential entailment notion as laid bare by Shoham ([Sho88]). This preferential entailment is different from Ruspini ([Rus91])'s approach to inference in fuzzy logic, which rather corresponds to an "approximate" entailment based on a similarity relation between interpretations, and not to a preference relation.

Possibilistic logic implements a non-monotonic reasoning in case of partial inconsistency. Indeed, it has been shown ([DP91b]) that the preferential entailment (in the sense of Shoham ([Sho88]) $\vDash_\pi$, defined as the inclusion of the preferred models of p (which maximize $\Pi(p)$) in the set of models of q, by

$$p \vDash_\pi q \Rightarrow \{\omega \mid \omega \in M(p), \pi(\omega) = \Pi(p) > 0\} \subseteq M(q) \Leftrightarrow N(q \mid p) > 0$$

where $$N(q \mid p) = 1 - \Pi(\neg q \mid p) \text{ and } \Pi(q \mid p) = \begin{cases} 1 \text{ if } \Pi(p) = \Pi(p \wedge q) \\ \Pi(p \wedge q) \text{ if } \Pi(p) > \Pi(p \wedge q) \end{cases}$$

and $\Pi(q \mid p)$ is the least specific solution of the equation $\Pi(p \wedge q) = \min(\Pi(q \mid p), \Pi(p))$ and where π is the possibility distribution associated with the semantics of the knowledge base $\mathcal{K}$, underlying Π. The entailment relation $\vDash_\pi$ is in complete agreement with non-monotonic consequence relations obeying the axiomatics of system P proposed by Kraus et al. ([KLM90]). For more details on possibilistic logic and its relation to symbolic approaches to nonmonotonic reasoning, see [DLP94a]).

As pointed out in [BDLP93], the weighted clause $(\neg p \vee q, \alpha)$, understood as $N(\neg p \vee q) \geq \alpha$ is semantically equivalent to the weighted clause $(q, \min(\alpha, v(p)))$ where v(p) is the truth value of p, i.e. v(p) = 1 if p is true and v(p) = 0 if p is false. This remark is very useful for hypothetical reasoning, since by "transferring" a sub-formula from a clause to the weight part of the formula we are introducing explicit assumptions. Indeed changing $(\neg p \vee q, \alpha)$ into $(q, \min(v(p), \alpha))$ leads to state the piece of knowledge under the form "q is certain at the degree α, *provided that* p is true". More generally, the weight or label can be a function of logical (universally quantified) variables involved in the clause. The weight is no more just a degree but in fact a label which expresses the context in which the piece of knowledge is more or less certain. This is to be related to "possibilistic Assumption-based Truth Maintenance Systems" (with weighted justifications and/or hypotheses, which have been defined and exemplified on a diagnosis problem ([BDLP93]). The approach contrasts with other uncertainty handling ATMS in the sense that the symbolic processing and the calculus of uncertainty are no longer separated here.

Moreover the presence of logical variables in the weight also enables the expression of some graduality attached to *vague predicates* (as in the certainty rule "the younger the person, the more certain he/she is single", where "young" is a vague predicate) in a simple way, as $N(\text{single}(x)) \geq \mu_{\text{young}}(\text{age}(x))$ in our example. It would then allow for a flexible interface between the symbolic knowledge base and numerical inputs. Vague predicates can thus be handled by introducing their characteristic functions in the weights; see [BDP92], [DLP94b].

3.5 Default Reasoning

The idea is to translate each default "if p then q, generally" denoted by $p \sim> q$ into a constraint expressing that the situation where p and q is true has a greater plausibility than the one where p and $\neg q$ is true; if we prefer, we want to express that in the context where p is true, q is more possible or plausible than $\neg q$. So we need a qualitative relation for comparing plausibility levels. We use a qualitative possibility relation $\geq_\Pi$ defined on a Boolean algebra of propositions ([Dub86]; [DP91d]; [Yag93]). Such a relation is assumed to be i) complete ($\forall p, \forall q, p \geq_\Pi q$ or $q \geq_\Pi p$), ii) transitive, and to satisfy iii) $\top >_\Pi \perp$, where $\top$ and $\perp$ represent tautology and contradiction respectively, and $>_\Pi$ the strict part of the ordering $\geq_\Pi$; iv) $\top \geq_\Pi p \geq_\Pi \perp$, $\forall p$; and the characteristic axiom v) $p \geq_\Pi q \Rightarrow r \vee p \geq_\Pi r \vee q$, $\forall p, \forall q, \forall r$. In the finite case, the only numerical counterparts to possibility relations are possibility measures, such that $\forall p, \forall q, \Pi(p \vee q) = \max(\Pi(p),\Pi(q))$ [Dub86].

A default $p \sim> q$ is then understood formally as the constraint

$$\Pi(p \wedge q) > \Pi(p \wedge \neg q)$$

on a possibility measure Π describing the semantics of the available knowledge. This constraint can be shown to be equivalent to $N(q \mid p) > 0$.

Let us consider the following (usual) set of default rules d_1: "birds fly", d_2: "penguins do not fly", d_3: "penguins are birds", symbolically written

$$d_1: b \sim> f$$
$$d_2: p \sim> \neg f$$
$$d_3: p \sim> b.$$

The set of three defaults will be represented by the following set C of constraints:

$$b \wedge f >_\Pi b \wedge \neg f$$
$$p \wedge \neg f >_\Pi p \wedge f$$
$$p \wedge b >_\Pi p \wedge \neg b.$$

Let Ω be the finite set of interpretations of our propositional language. If this language is made of the literals $p_1, \ldots, p_n$, these interpretations correspond to the possible worlds where the conjunctions $*p_1 \wedge \ldots \wedge *p_n$ are true, where $*$ stands for the presence of the negation sign $\neg$ or its absence. In our example, $\Omega = \{\omega_0: \neg b \wedge \neg f \wedge \neg p, \omega_1: \neg b \wedge \neg f \wedge p, \omega_2: \neg b \wedge f \wedge \neg p, \omega_3: \neg b \wedge f \wedge p, \omega_4: b \wedge \neg f \wedge \neg p, \omega_5: b \wedge \neg f \wedge p, \omega_6: b \wedge f \wedge \neg p, \omega_7: b \wedge f \wedge p\}$. Then the set of constraints C' on models is:

$$C'_1: \max(\pi(\omega_6), \pi(\omega_7)) > \max(\pi(\omega_4), \pi(\omega_5))$$
$$C'_2: \max(\pi(\omega_5), \pi(\omega_1)) > \max(\pi(\omega_3), \pi(\omega_7))$$
$$C'_3: \max(\pi(\omega_5), \pi(\omega_7)) > \max(\pi(\omega_1), \pi(\omega_3)).$$

Let $>_\pi$ be a ranking of Ω, such that $\omega >_\pi \omega'$ iff $\pi(\omega) > \pi(\omega')$ on Ω. Any finite consistent set of constraints: $p_i \wedge q_i >_\Pi p_i \wedge \neg q_i$ induces a partially defined ranking $>_\pi$ on Ω, that can be completed according to the principle of minimum specificity. The idea is to try to assign to each world ω the highest possibility level (in forming a well-ordered partition of Ω) without violating the constraints. The ordered partition of Ω

associated with $>_\pi$ using the minimum specificity principle can be easily obtained by the following procedure ([BDP92])

a. $i = 0$
b. While Ω is not empty repeat b.1.-b.4.:
   b.1. $i \leftarrow i + 1$
   b.2. Put in E_i every model which does not appear in the right side of any constraints of C',
   b.3. Remove the elements of E_i from Ω,
   b.4. Remove from C' any constraint containing elements of E_i.

Let us apply now this algorithm. The models which do not appear in the right side of any constraint of C' are $\{\omega_0, \omega_2, \omega_6\}$, we call this set E_1. We remove the elements of E_1 from Ω and we remove the constraint C'_1 from C' (since assigning to ω_6 the highest possibility level makes the constraint C'_1 always satisfied). We start again the procedure and we find successively the two following sets $\{\omega_4, \omega_5\}$ and $\{\omega_1, \omega_3, \omega_7\}$. Finally, the well ordered partition of Ω is:

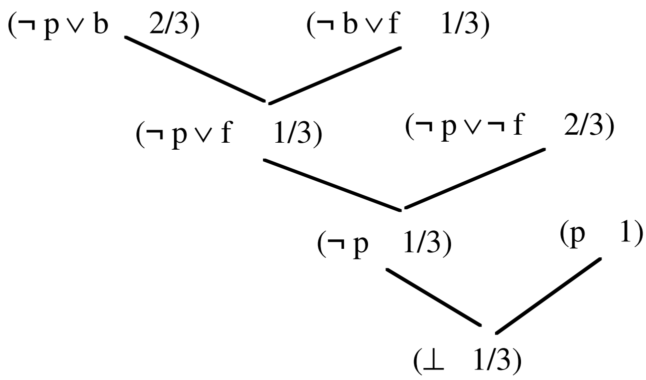
$$\{\omega_0, \omega_2, \omega_6\} >_\pi \{\omega_4, \omega_5\} >_\pi \{\omega_1, \omega_3, \omega_7\}.$$

Let $E_1, \ldots, E_m$ be the obtained partition. A numerical counterpart to $>_\pi$ can be defined by

$$\pi(\omega) = \frac{m + 1 - i}{m} \text{ if } \omega \in E_i,\ i = 1,m.$$

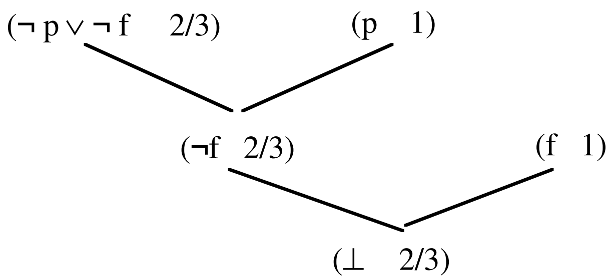
In our example we have $m = 3$ and $\pi(\omega_0) = \pi(\omega_2) = \pi(\omega_6) = 1$; $\pi(\omega_4) = \pi(\omega_5) = 2/3$; $\pi(\omega_1) = \pi(\omega_3) = \pi(\omega_7) = 1/3$. Note that this is purely a matter of convenience to use a numerical scale, and any other numerical counterpart such that $\pi(\omega) > \pi(\omega')$ iff $\omega >_\pi \omega'$ will work as well. Namely π is used as an ordinal scale. From this possibility distribution, we can compute for any proposition p its necessity degree N(p). For instance, $N(\neg p \vee \neg f) = \min\{1 - \pi(\omega) \mid \omega \models p \wedge f\} = \min(1 - \pi(\omega_3), 1 - \pi(\omega_7)) = 2/3$, while $N(\neg b \vee f) = \min\{1 - \pi(\omega) \mid \omega \models b \wedge \neg f\} = \min(1 - \pi(\omega_4), 1 - \pi(\omega_5)) = 1/3$ and $N(\neg p \vee b) = \min(1 - \pi(\omega_1), 1 - \pi(\omega_3)) = 2/3$.

The method then consists in turning each default $p_i \rightarrow q_i$ into a possibilistic clause $(\neg p_i \vee q_i, N(\neg p_i \vee q_i))$ where N is computed from the possibility distribution π induced by the set of constraints corresponding to the default knowledge base. Then we apply the possibilistic inference machinery for reasoning with the defaults together with the available factual knowledge.



The possibilistic knowledge base equivalent to the set of defaults we have is $\Sigma = \{(\neg b \vee f \;\; 1/3), (\neg p \vee b \;\; 2/3), (\neg p \vee \neg f \;\; 2/3)\}$. Then the following derivation, knowing with certainty that Tweety is a penguin (p 1) and a bird (b 1), gives the optimal degree of inconsistency of $\Sigma \cup \{(p\ 1), (b\ 1)\}$ which is equal to 1/3 by applying the resolution rule of possibilistic logic repeatedly:

By refutation, the following derivation shows that ¬f is truly a logical consequence of $\Sigma \cup \{(p\ 1), (b\ 1)\}$, i.e., Tweety does not fly since adding the piece of information (f 1) we find the degree of inconsistency equal to 2/3 which is higher than 1/3:



This simple example suggests how possibility theory can cope with default reasoning is a simple and rather elegant way.

3.6 Abductive Reasoning

This section points out the interest of possibility and necessity measures for describing plausible links between disorders and manifestations in a diagnosis problem. The presentation is made in terms of sets of manifestations and of relations between disorders and manifestations, rather than in logical terms, for simplicity.

Let $\mathcal{S}$ be a system whose current state is described by means of a n-tuple of *binary* attributes $(a_1, \ldots, a_i, \ldots, a_n)$. When $a_i = 1$ we shall say that the manifestation m_i is present; when $a_i = 0$, it means that m_i is absent. When there is no manifestation present, $\mathcal{S}$ is said to be in its normal state and its state is described by the n-tuple $(0, \ldots, 0, \ldots, 0)$. Let $\mathcal{M}$ denote the set of the n possible manifestations $\{m_1, \ldots, m_i, \ldots, m_n\}$. Let $\mathcal{D}$ be a set of possible disorders $\{d_1, \ldots, d_j, \ldots, d_k\}$. A disorder can be present or absent. We assume here that its presence is not a matter of degree. To each d_i we associate the set $M(d_j)$ of manifestations which are entailed, or if we prefer caused, produced, by the presence of d_j *alone*. We first consider the completely informed case where all the present manifestations are observed and where the set of manifestations which appear when a disorder is present is perfectly known. Thus if $m_i \notin M(d_j)$ it means that m_i cannot be caused by d_j. We thus define a relation R on $\mathcal{D} \times \mathcal{M}$, defined by $(d_j,m_i) \in R \Leftrightarrow m_i \in M(d_j)$, which associates manifestations and disorders. Such a relational approach has been particularly developed by Peng and Reggia [PR90].

Given a set M^+ of present manifestations which are observed, the problem is to find what disorder(s) *may* have produced the manifestations in M^+. We suppose that the set $M^- = \mathcal{M} - M^+ = \overline{M^+}$ is the set of manifestations which are absent, i.e. all manifestations which are present are observed. While deductive reasoning enables us to predict the presence of manifestation(s) from the presence of disorder(s), abductive reasoning looks for possible cause(s) of observed effects. In other words, we look for plausible explanations (in terms of disorders) of an observed situation. Clearly while it is at least theoretically possible to find out all the possible causes which may have led to a given state of the system $\mathcal{S}$, the ordering of the possible solutions according to some levels of plausibility is out of the scope of logical reasoning, strictly speaking. However, one may consider that an explanation of a set of manifestations is more reasonable when it involves a small number of disorders being conjointly present: a single cause is more acceptable than several jointly present ones. This is called the principle of parsimony. For the case of a set of disorders which jointly causes the manifestations the reader is referred to [DP93a]. In the completely informed case described above, we have i) $M^+ = \overline{M^-}$, i.e., all the present manifestations are observed, and equivalently all the manifestations which are not observed are indeed absent, and ii) $\forall d$, $M(d) = M(d)^+ = \overline{M(d)^-}$, where $M(d)^+$ (resp. $M(d)^-$) is the set of manifestations which are certainly present (resp. certainly absent) when disorder d alone is present. In the general case, we only have $M^+ \cap M^- = \emptyset$ but $M^+ \cup M^- \neq \mathcal{M}$ and similarly $M(d)^+ \cap M(d)^- = \emptyset$, $M(d)^+ \cup M(d)^- \neq \mathcal{M}$, $\forall d$. Let $M° = \mathcal{M} - M^+ - M^-$.

Let us consider the general case where both the information pertaining to the manifestations and the information relative to the association between disorders and manifestations is incomplete. Then d belongs to the set $\widehat{D}$ of potential disorders each of which can alone explain both M^+ and M^- if and only if d does not produce with certainty any manifestation which is certainly absent in the evidence, and no observed manifestations is ruled out by d. Formally we have

$$\widehat{D} = \{d \in \mathcal{D},\ M(d)^+ \subseteq \overline{M^-} \text{ and } M(d)^- \subseteq \overline{M^+}\}. \tag{40}$$

This also writes

$$\widehat{D} = \{d \in \mathcal{D},\ M(d)^+ \cap M^- = \emptyset \text{ and } M(d)^- \cap M^+ = \emptyset\}. \tag{41}$$

$\widehat{D}$ gathers all the disorders in $\mathcal{D}$ which cannot be ruled out from the observations. If $M^- = \emptyset$ and $M(d)^- = \emptyset$, i.e. no information is available on the manifestations certainly absent, it can be verified that $\widehat{D} = \mathcal{D}$ and the whole set of possible disorders is obtained.

The unknown disorder d_0 always belongs to $\widehat{D}$. The membership test (40) is thus very permissive, i.e., $\widehat{D}$ can be very large, and contains what may look like irrelevant causes, when they cannot be ruled out. Namely a disorder d may belong to $\widehat{D}$ defined by (40)-(41) even if $M(d)^+ \cap M^+ = \emptyset$ and $M(d)^- \cap M^- = \emptyset$. Indeed, in this case, $M(d)^+ \cup M(d)^- \subseteq M^0$. This means that $\widehat{D}$ includes disorders none sure manifestations of which are observed and no forbidden manifestations are for sure absent. Such a disorder d may still be present when M^+ and M^- are observed, since all the available, certain information we have about d pertains to manifestations in M^0 about which no observation exists (this is true for the unknown disorder d_0). Then, among the disorders in $\widehat{D}$ we may prefer those, if any, which have at least a weak relevance to the observations, namely the subset $\widehat{D}^*$ of $\widehat{D}$, defined by

$$\widehat{D}^* = \{d \in \widehat{D}, M(d)^+ \cap M^+ \neq \emptyset \text{ or } M(d)^- \cap M^- \neq \emptyset\}. \tag{42}$$

This eliminates the disorders such that $M(d)^+ \cup M(d)^- \subseteq M^0$, i.e., the disorders which are not suggested by the observations without being ruled out by them (and d_0 particularly).

In this section, we present a graded extension of the above model. Namely, M^+ and M^- are now fuzzy sets of manifestations which are more or less certainly present, and more or less certainly absent respectively. However we keep the requirement $M^+ \cap M^- = \emptyset$ (where the intersection is defined by the min operation), i.e. we cannot be somewhat certain both of the presence and of the absence of the same manifestation simultaneously. Similarly, $M(d)^+$ and $M(d)^-$ will denote the fuzzy sets of manifestations which are respectively more or less certainly present and more or less certainly absent when disorder d alone is present (more generally when the subset D of disorders is present). Obviously, we also assume $\forall d$, $M(d)^+ \cap M(d)^- = \emptyset$.

By complementation (defined by $\mu_{\overline{F}} = 1 - \mu_F$) of M^- and $M(d)^-$, we obtain the fuzzy sets $\overline{M^-}$, $\overline{M(d)^-}$ of manifestations which are more or less possibly present, respectively, in the considered situation, when d is present. This corresponds to the usual duality between what is (more or less) certain, i.e. necessarily true, and what is (more or less) possibly true. Indeed a pair of dual possibility and necessity measures Π and N are related by the relation $\Pi(A) = 1 - N(\bar{A})$, for any event A (here A represents the presence of a manifestation).

Note that $M^+ \subseteq \overline{M^-}$, $M(d)^+ \subseteq \overline{M(d)^-}$, in the sense of fuzzy set inclusion. An even stronger inclusion holds. Since $M^+ \cap M^- = \emptyset$, we have

$$\{m_i \in \mathcal{M}, \mu_{M^+}(m_i) > 0\} \subseteq \{m_i \in \mathcal{M}, \mu_{M^-}(m_i) = 0\} = \{m_i \in \mathcal{M}, \mu_{\overline{M^-}}(m_i) = 1\} \tag{43}$$

i.e. the support of M^+ (i.e. the elements with a positive membership grade) is included in the core of $\overline{M^-}$ (the elements which belong with degree 1); the same holds for $M(d)^+$, $\overline{M(d)^-}$. This is in agreement with the fact that for crisp events A, we have

$N(A) > 0 \Leftrightarrow \Pi(\bar{A}) < 1 \Rightarrow \Pi(A) = 1$ since then one of A or $\bar{A}$, at least, should be completely possible. A pair of fuzzy sets (F,G) such that $F \cap \bar{G} = \emptyset$ is called a twofold fuzzy set [DP87b]. Twofold fuzzy sets (F,G) have been introduced for modelling incompletely known sets, i.e. sets for which we know elements gathered in F, which more or less certainly belong to it, as well as other elements, gathered in $\bar{G}$, which more or less certainly do not belong to it. But $F \cup \bar{G}$ may not cover the whole referential. Similarly, the pair $(M(d)^+, \overline{M(d)^-})$ defines twofold fuzzy relations on $\mathcal{D} \times \mathcal{M}$.

The extension to fuzzy sets of equations (40) and (41) can be done very simply. It requires that the extent to which two fuzzy sets F and G of $\mathcal{M}$ intersect be evaluated. The consistency between F and G is simply defined as [Zad79a]

$$\mathrm{cons}(F,G) = \sup_{m \in \mathcal{M}} \min(\mu_F(m), \mu_G(m)). \tag{44}$$

It computes the degree of existence of some common element for F and G. (41) is based on checking the inconsistency level between fuzzy sets, that is 1 – cons(F,G). The fuzzy extension of (41) then leads to compute a fuzzy set $\widehat{D}$ of potential unique disorders as,

$$\begin{aligned} \forall\, d \in \mathcal{D},\ \mu_{\widehat{D}}(d) &= \min(1 - \mathrm{cons}(M(d)^+, M^-), 1 - \mathrm{cons}(M(d)^-, M^+)) \\ &= 1 - \max(\mathrm{cons}(M(d)^+, M^-), \mathrm{cons}(M(d)^-, M^+)) \end{aligned} \tag{45}$$

.where the minimum operator expresses the conjunction of the conditions in (41). It is easy to check that (43) can be written in terms of inclusion indices, like (40), under form

$$\mu_{\widehat{D}}(d) = \min(\mathrm{inc}(M(d)^+, \overline{M^-}), \mathrm{inc}(M(d)^-, \overline{M^+})) \tag{46}$$

where the inclusion of fuzzy set F into fuzzy set G is defined by

$$\mathrm{inc}(F,G) = 1 - \mathrm{cons}(F,\bar{G}) = \inf_{m \in \mathcal{M}} \max(1 - \mu_F(m), \mu_G(m)). \tag{47}$$

Note that $\mathrm{inc}(F,G) = 1 \Leftrightarrow \mathrm{support}(F) \subseteq \mathrm{core}(G) \Leftrightarrow F \cap \bar{G} = \emptyset$, i.e. the inclusion index extends the strong inclusion encountered in (43) and guarantees that d fully belongs to $\widehat{D}$ (i.e. with degree 1) if and only if $M(d)^+$ and $\overline{M^-}$, $M(d)^-$ and $\overline{M^+}$ do not overlap at all. (45) clearly expresses that a disorder d is all the less a candidate explanation as the fuzzy set of its more or less certain effects overlaps the fuzzy set of manifestations more or less certainly absent, or as the fuzzy set of effects which are more or less certainly absent when d is present overlaps the fuzzy set of manifestations which are more or less certainly present. This is intuitively satisfying.

In the particular case where $M^- = \emptyset$, i.e., we only have positive observations, $\widehat{D}$ is defined by

$$\forall\ d \in \mathcal{D},\ \mu_{\widehat{D}}(d) = \min_{i=1,n} \max(1-\mu_{M^+}(m_i), \mu_{\overline{M(d)}^-}(m_i)). \tag{48}$$

This case corresponds to the situation studied by Peng and Reggia [PR90] and $\mu_{\widehat{D}}(d)$ evaluates to what extent d covers M^+. The idea of relevance expressed by equation (42) in the crisp case can be easily extended using the consistency index. Namely,

$$\mu_{\widehat{D}^*}(d) = \min(\mu_{\widehat{D}}(d), \max(\mathrm{cons}(M(d)^+, M^+), \mathrm{cons}(M(d)^-, M^-))). \tag{49}$$

Note that the modelling of uncertainty remains qualitative in this approach as in the previous section. Indeed, we could use a finite completely ordered chain of levels of certainty ranging between 0 and 1, i.e. $\ell_1 = 0 < \ell_2 < \ldots < \ell_n = 1$ instead of [0,1], with $\min(\ell_i, \ell_k) = \ell_i$ and $\max(\ell_i, \ell_k) = \ell_k$ if $i \leq k$, and $1 - \ell_i = \ell_{n+1-i}$.

Taking into account the incomplete nature of the information about the presence or absence of manifestations decreases the discrimination power when going from the completely informed case to the incomplete information case, since then the number of possible disorders in $\widehat{D}$ increases. This is due to the fact that now there are manifestations which are neither certain nor impossible and consequences of the presence of a given disorder d which are only possible, as pictured in Figure 5, while $M(d)^+ = M(d)$ and $M(d)^- = \overline{M(d)}$ in Figure 4 (similar figures can be drawn for M^+ and M^-).

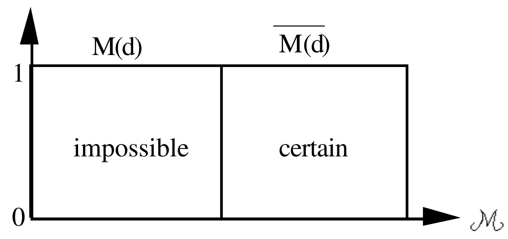


Figure 4. Completely informed situation

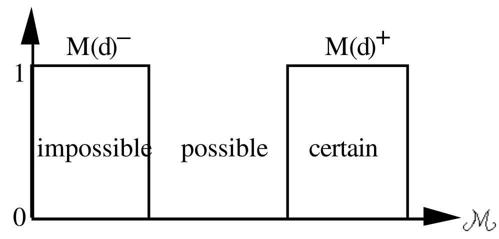


Figure 5. Partially informed situation

This suggests that, in order to improve the discrimination power of the model, we have to refine the non-fuzzy model in such a way that consequences (resp. manifestations) previously expressed as certain (resp. certainly present) and impossible (resp. certainly absent) remain classified in the same way and where some possible consequences (resp. possibly present manifestations) are now allowed to be either somewhat certain (resp. somewhat certainly present) or somewhat impossible (resp. somewhat certainly absent). See Figure 6. Then (46) enables us to rank-order the possible disorders which are compatible with the observations. This counterbalances the increase of candidates due to the incompleteness of the information.

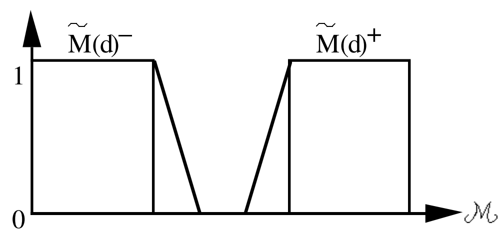


Figure 6. Partially informed situation with preferences

It can be easily shown that adding preference levels on top of incompleteness modeling can at the same time enable the discrimination power of the completely informed situation to be recovered, and also enable the extra disorders obtained on the partially informed situation to be ranked in terms of their plausibility levels. See [DP93a].

Example: Let us show the benefit of grading uncertainty on an example. Let $\mathcal{D}$ = $\{d_1, d_2, d_3, d_4\}$, $\mathcal{M}$ = $\{m_1, m_2, m_3\}$ and let us first consider the completely informed case

$$M(d_1)^+ = \{m_2\},\ M(d_1)^- = \{m_1, m_3\}$$
$$M(d_2)^+ = \{m_3\},\ M(d_2)^- = \{m_1, m_2\}$$
$$M(d_3)^+ = \{m_1, m_2, m_3\},\ M(d_3)^- = \emptyset$$
$$M(d_4)^+ = \{m_2, m_3\},\ M(d_4)^- = \{m_1\},$$

and $$M^+ = \{m_2, m_3\},\ M^- = \{m_1\}.$$

Then, we get $\widehat{D} = \{d_4\}$.

Let us weaken our example into an incompletely informed case

$$M(d_1)^+ = \emptyset,\ M(d_1)^- = \{m_1\}$$
$$M(d_2)^+ = \{m_3\},\ M(d_2)^- = \emptyset$$
$$M(d_3)^+ = \{m_3\},\ M(d_3)^- = \emptyset$$
$$M(d_4)^+ = \emptyset,\ M(d_4)^- = \{m_1\}.$$

Then $\widehat{D} = \{d_1, d_2, d_3, d_4\}$, i.e. all causes are possible. Let us now consider the following table where the previous incompletely informed case is completed with some gradual uncertainty information that is in accordance with the complete information case.

	$M(d_1)^+$	$M(d_1)^-$	$M(d_1)^+$	$M(d_2)^-$	$M(d_3)^+$	$M(d_3)^-$	$M(d_4)^+$	$M(d_4)^-$	$\mathcal{M}$	
membership m_1	0	1	0	0.9	0.7	0	0	1	0	M^-
m_2	0.8	0	0	0.6	0.5	0	0.7	0	1	M^+
grades m_3	0	0.8	1	0	1	0	0.3	0	1	M^+
m_3	0	0.8	1	0	1	0	0.3	0	1	M^+
$cons(M(d)^+, M^-)$	0		0		0.7		0			
$cons(M(d)^+, M^-)$		0.8		0.6		0		0		
form. (45): 1 – max	0.2		0.4		0.3		1			

Thus we obtain that d_4 is much more plausible than d_2, itself more plausible than d_3, which is more plausible than d_1, i.e. a ranking of the possible disorders is obtained. This is in accordance with the complete information case as well as the incomplete information case, but significantly improves their findings. Note that all disorders are relevant in the sense of (49). The reader is referred to [CDHP94]) for the application of this approach to a satellite diagnosis problem.

4 CONCLUDING REMARKS

This overview paper has provided a description of some fuzzy set or possibility theory-based methods in approximate reasoning which correspond to two interpretations of the idea of a fuzzy as either describing a gradual property (or if we prefer a fuzzy predicate) or modelling an incomplete state of information. This paper offers only a partial overview of the existing works in approximate reasoning, and is centered on the work done by the authors for about 10 years. There exists a third interpretation which has not been presented here for the sake of brevity, but which is also worth-considering. This is when a fuzzy set models a flexible constraint expressing what are the values which are allowed to be chosen for a given quantity or parameter. In such a situation the semantics of the possibility distribution is in terms of preference. This has not to be confused with a possibility distribution representing an incomplete state of knowledge. Indeed in that latter case the ill-known value restricted by the possibility distribution is not a matter of choice, but rather of randomness. The extension of Constraint Satisfaction Problem methods to flexible constraints modelled in the fuzzy set framework has been recently thoroughly investigated; see [DFP94]).

Besides, the paper has advocated possibility theory as a convenient framework for handling uncertainty in a rather qualitative way, but no comparison has been provided with probability-based methods. The reader interested in a comparative view and a discussion on the differences but also on the links between possibilistic and the probabilistic frameworks are referred to [Bez94] and [DP93b].

REFERENCES

[Bal79] J.F. Baldwin. A new approach to approximate reasoning using a fuzzy logic. *Fuzzy Sets and Systems*, 2:309-325, 1979.

[BDLP93] S. Benferhat, D. Dubois, J. Lang, and H. Prade. Hypothetical reasoning in possibilistic logic: Basic notions, applications and implementation issues. In P.Z. Wang and K.F. Loe, editors. *Between Mind and*

Computer — Fuzzy Science and Engineering, pages 1-29. World Scientific, Singapore, 1993.

[BDP92] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In B. Nebel, C. Rich and W. Swartout, editors. *Proc. of the 3rd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 673-684, Cambridge, MA, Oct. 25-29, 1992.

[Bez94] J.C. Bezdek (Eds.). *Special Issue "Fuzziness vs. Probability — The n-th round*. IEEE Trans. on Fuzzy Systems, 2(1): 1-42, 1994.

[Bou87] B. Bouchon. Fuzzy inferences and conditional possibility distributions. *Fuzzy Sets and Systems*, 23:33-41, 1987.

[Dub86] D. Dubois. Belief structures, possibility theory and decomposable confidence measures on finite sets. *Comput. Artif. Intell.* (Bratislava), 5(5):403-416, 1986.

[CDHP94] D. Cayrac, D. Dubois, M. Haziza, and H. Prade. Possibility theory in 'fault mode effect analyses' —A satellite fault diagnosis application—. In *Proc. of the 3rd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'94)*, Orlando, June 1994.

[DFP94] D. Dubois, H. Fargier, and H. Prade. Propagation and satisfaction of flexible constraints. In R.R. Yager and L.A. Zadeh, editors. *Fuzzy Sets, Neural Networks and Soft Computing*, pages 166-187. Van Nostrand, 1994.

[DGP94] D. Dubois, M. Grabisch, and H. Prade. Gradual rules and the approximation of control laws. To appear in H.T. Nguyen, M. Sugeno, R. Tong and R.R. Yager, editors. *Theoretical Aspects of Fuzzy Control*. Wiley, New York, 1994.

[DLP92] D. Dubois, J. Lang, and H. Prade. Advances in automated reasoning using possibilistic logic. In A. Kandel, editor. *Fuzzy Expert Systems*, pages 12-134. CRC Press, Boca Raton, FL, 1992.

[DLP94a] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D.M. Gabbay, C.J. Hogger, J.A. Robinson and D. Nute, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 439-513. Clarendon Press, Oxford, 1994.

[DLP94b] D. Dubois, J. Lang, and H. Prade. Automated reasoning using possibilistic logic: semantics, belief revision, and variable certainty weights. *IEEE Trans. on Knowledge and Data Engineering*, 6:64-71, 1994.

[DMP88] D. Dubois, R. Martin-Clouaire, and H. Prade. Practical computing in fuzzy logic. In M.M. Gupta and T. Yamakawa, editors. *Fuzzy Computing — Theory, Hardware, and Applications*, pages 11-34. North-Holland, Amsterdam, 1988.

[DP80] D. Dubois, and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[DP87a] D. Dubois, and H. Prade. Fuzzy numbers: An overview. In J.C. Bezdek, editor. *Analysis of Fuzzy Information — Vol. I: Mathematics and Logic*, pages 3-39. CRC Press, Boca Raton, FL, 1987.

[DP87b] D. Dubois, and H. Prade. Twofold fuzzy sets and rough sets — Some issues in knowledge representation. *Fuzzy Sets and Systems*, 23:3-18, 1987.

[DP88] 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*. Plenum Press, New York, 1988.

[DP89] D. Dubois, and H. Prade. Order-of-magnitude reasoning with fuzzy relations. *Revue d'Intelligence Artificielle* (Hermès, Paris), 3:69-94, 1989.

[DP91a] D. Dubois, and H. Prade. Semantic considerations on order of magnitude reasoning. In M.G. Singh and L. Travé-Massuyès, editors. *Decision Support Systems and Qualitative Reasoning*, pages 223-228. North-Holland, Amsterdam, 1991.

[DP91b] D. Dubois, and H. Prade. Possibilistic logic, preference models, non-monotonicity and related issues. In *Proc. of the 12th Inter. Joint Conf. on Artificial Intelligence (IJCAI'91)*, pages 419-424, Sydney, Australia, Aug. 24-30, 1991.

[DP91c] D. Dubois, and H. Prade. Fuzzy sets in approximate reasoning. *Fuzzy Sets and Systems*, Part 1: 40:143-202; Part 2 (with Lang J.): 40:203-244, 1991.

[DP91d] D. Dubois, and H. Prade. Epistemic entrenchment and possibilistic logic. *Artificial Intelligence,* 50:223-229, 1991

[DP92a] D. Dubois, and H. Prade. Possibility theory as a basis for preference propagation in automated reasoning. In *Proc. of the 1st IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'92)*, pages 821-832, San Diego, CA, March 8-12, 1992.

[DP92b] D. Dubois, and H. Prade. Fuzzy rules in knowledge-based systems — Modelling gradedness, uncertainty and preference. In R.R. Yager and L.A. Zadeh, editors. *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, pages 45-68. Kluwer Academic Publ., Dordrecht, 1992.

[DP93a] D. Dubois, and H. Prade. A fuzzy relation-based extension of Reggia's relational model for diagnosis handling uncertain and incomplete information. In D. Heckerman and A. Mamdani, editors. *Proc. of the 9th Conf. on Uncertainty in Artificial Intelligence*, pages 106-113, 1993.

[DP93b] D. Dubois, and H. Prade. Fuzzy sets and probability: Misunderstandings, bridges and gaps. In *Proc. of the 2nd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'93)*, pages 1059-1068, San Francisco, CA, March 28-April 1st, 1993.

[DSPS89] A. Di Nola, S. Sessa, W. Pedrycz, and E. Sanchez. *Fuzzy Relation Equations and Their Applications to Knowledge Engineering*. Kluwer Academic Publ., Dordrecht, 1989.

[KF88] G.J. Klir, and T.A. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[KK93a] F. Klawonn, and R. Kruse. Equality relations as a basis for fuzzy control. *Fuzzy Sets and Systems*, 54:147-156, 1993.

[KK93b] F. Klawonn, and R. Kruse. Fuzzy control as interpolation on the basis of equality relations. In *Proc. of the 2nd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'93)*, pages 1125-1130, San Francisco, CA, 1993.

[KLM90] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:134-207, 1990.

[LDP91] J. Lang, D. Dubois, and H. Prade. A logic of graded possibility and certainty coping with partial inconsistency. In *Proc. of the 7th Conf. on Uncertainty in Artificial Intelligence*, pages 188-196, Los Angeles, CA, July 13-15, 1991.

[Mam77] E.H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Trans. Comput.*, 26:1182-1191, 1977.

[NR75] C.V. Negoita, and D.A. Ralescu. *Applications of Fuzzy Sets to Systems Analysis*. ISR 11, Birkhaeuser, Basel, 1975.

[Ovc91] S.V. Ovchinnikov. Similarity relations, fuzzy partitions, and fuzzy orderings. *Fuzzy Sets and Systems*, 40:107-126, 1991.

[PR90] Y. Peng, and J.A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer Verlag, Berlin, 1990.

[Rus91] E.H. Ruspini. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning*, 5:45-88, 1991.

[San78] E. Sanchez. On possibility-qualification in natural languages. *Information Sciences*, 15:45-76, 1978.

[Sho88] Y. Shoham. *Reasoning About Change — Time and Causation from the Standpoint of Artificial Intelligence*. The MIT Press, Cambridge, MA, 1988.

[SN85] M. Sugeno, and M. Nishida. Fuzzy control of model car. *Fuzzy Sets and Systems*, 16:103-113, 1985.

[SS83] B. Schweizer, and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.

[Sug85] M. Sugeno. An introductory survey of fuzzy control. *Information Sciences*, 36:59-83, 1985.

[Tsu79] T. Tsukamoto. An approach to fuzzy reasoning method. In M.M. Gupta, R.K. Ragade and R.R. Yager, editors. *Advances in Fuzzy Set Theory and Applications*, pages 137-149. North-Holland, Amsterdam, 1979.

[TV85] E. Trillas, and L. Valverde. On mode and implication in approximate reasoning. In M.M. Gupta, A. Kandel, W. Bandler, J.B. Kiszka, editors. *Approximate Reasoning in Expert Systems*, pages 157-166. North-Holland, Amsterdam, 1985.

[Yag83] R.R. Yager. An introduction to applications of possibility theory. *Human Systems Management*, 3:246-269, 1983.

[Yag93] R.R. Yager. On the completion of qualitative possibility measures. *IEEE Trans. on Fuzzy Systems*, 1(3):184-194, 1993.

[Zad65] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[Zad71] L.A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 177-200, 1971.

[Zad73] L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man and Cybernetics*, 2:28-44, 1973.

[Zad75] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, Part 1: 8:199-249; Part 2: 2:301-357; Part 3: 9:43-80, 1975.

[Zad78a] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3-28, 1978.

[Zad78b] L.A. Zadeh. PRUF – A meaning representation language for natural languages. *Int. J. of Man-Machine Studies*, 10:395-460, 1978.

[Zad79a] L.A. Zadeh. A theory of approximate reasoning. In J.E. Hayes, D. Michie and L.I. Mikulich, editors. *Machine Intelligence*, Vol. 9, pages 149-194. Elsevier, Amsterdam, 1979.

[Zad79b] L.A. Zadeh. Fuzzy sets and information granularity. In M.M. Gupta, R.K. Ragade and R.R. Yager, editors. *Advances in Fuzzy Set Theory and Applications*, pages 3-18. North-Holland, Amsterdam, 1979.