Fuzzy sets and probability : Misunderstandings, bridges and gaps

Didier Dubois – Henri Prade
Institut de Recherche en Informatique de Toulouse (I.R.I.T.) – C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne
31062 Toulouse Cedex, France

INTRODUCTION

One of the most controversial issues in uncertainty modelling and information sciences is the relationship between probability theory and fuzzy sets. This paper is meant to survey the literature pertaining to this debate, and to try to overcome misunderstandings and to supply access to many basic references that have addressed the "probability versus fuzzy set" challenge. This problem has not a single facet, as will be claimed here. Moreover it seems that a lot of controversies might have been avoided if protagonists had been patient enough to build a common language and to share their scientific backgrounds. The main points made here are as follows. i) Fuzzy set theory is a consistent body of mathematical tools. ii) Although fuzzy sets and probability measures are distinct, several bridges relating them have been proposed that should reconcile opposite points of view ; especially possibility theory stands at the cross-roads between fuzzy sets and probability theory. iii) Mathematical objects that behave like fuzzy sets exist in probability theory. It does not mean that fuzziness is reducible to randomness. Indeed iv) there are ways of approaching fuzzy sets and possibility theory that owe nothing to probability theory. Interpretations of probability theory are multiple especially frequentist versus subjectivist views (Fine [31]) ; several interpretations of fuzzy sets also exist. Some interpretations of fuzzy sets are in agreement with probability calculus and some are not.

The paper is structured as follows : first we address some classical misunderstandings between fuzzy sets and probabilities. They must be solved before any discussion can take place. Then we consider probabilistic interpretations of membership functions, that may help in membership function assessment. We also point out non-probabilistic interpretations of fuzzy sets. The next section examines the literature on possibility-probability transformations and tries to clarify some lurking controversies on that topic. In conclusion, we briefly mention several subfields of fuzzy set research where fuzzy sets and probability are conjointly used.

1. MISUNDERSTANDINGS

1.1. Membership Function and Probability Measure

At the formal level, the relative situation of fuzzy sets and probability is clear. However some risk for confusion are already present in the way the definitions are understood. After Zadeh [81] a fuzzy set F on a universe U is defined by a membership function $\mu_F : U \to [0,1]$ and $\mu_F(u)$ is the grade of membership of *element u* in F. Here let us restrict to finite universes for simplicity. In contrast, a probability measure P is a mapping $2^U \to [0,1]$ that assigns a number P(A) to each *subset* of U, and satisfies the Kolmogorov axioms :

$$\begin{aligned} & P(U) = 1 \ ; \ P(\emptyset) = 0 \\ \text{if } A \cap B = \emptyset \quad & P(A \cup B) = P(A) + P(B) \end{aligned} \tag{1}$$

P(A) is the probability that an ill-known single-valued variable x ranging on U hits the fixed well-known set A. A first misunderstanding is to confuse P(A) with a membership grade. But when we consider $\mu_F(u)$, the element u is fixed and known and the set is ill-defined, while when we consider P(A), the set A is well defined while the value of the underlying variable x, to which P is attached, is unknown (and moves). $\mu_F(u)$ has more similarity with $P(\{u\}) = p(u)$, except that the condition $\sum_{u\in U} p(u) = 1$ must hold while it is not so for μ_F. A set-theoretic calculus for probability distributions has been recently built up under the name of Lebesgue logic (Bennett et al. [5]).

At the mathematical level the domain of the mapping P is the Boolean algebra 2^U while the set of fuzzy sets is $[0,1]^U$, i.e. can never be a Boolean algebra. It has necessarily a weaker structure, depending on the choice of fuzzy set operations. One should not assimilate (e.g. Cooke [9]) the definition of a fuzzy set connective, e.g. union

$$\mu_{F\cup G}(u) = S(\mu_F(u),\mu_G(u)) \tag{2}$$

with the decomposability property

$$P(A \cup B) = S'(P(A),P(B)) \tag{3}$$

of the probability measure with respect to two given events A and B. First (3) cannot hold for all A and B when P is a probability measure while (2) can be assumed for all $u \in U$. The temptation to get confused lies in the fact (due to the Boolean algebra structure of 2^U) that S and S' can both be choosen among a family of semi-groups of the unit interval (i.e. triangular co-norms such as max(a,b), a + b – ab, or min(1, a + b)). This is because the properties usually required for membership grade combinations (e.g. the associativity of union and the empty set acting as an identity leads to choose S among triangular co-norms), and the properties induced by the structure of 2^U on degrees of probability are the same (e.g. Dubois and Prade [16]). Concludingly, using min and max is not enough to claim that fuzzy sets are at work ; similarly the fact that sometimes $P(A \cup B) = \max(P(A),P(B))$ (e.g. if $A \subseteq B$) is not enough to claim that fuzzy sets are probabilities. This type of confusion between degrees of probability and membership grades pervades the literature of uncertainty in expert systems (see Duda et al. [23], Horvitz et al. [42] for instance).

A dual point of view is to start with a poorly structured set of sentences $\mathcal{S}$ and consider various ways of attaching numbers to sentence $\alpha \in \mathcal{S}$, say $v(\alpha) \in [0,1]$. The question is then to analyze how the structure of $\mathcal{S}$ under the logical connectives AND, OR, NOT, etc… is mapped on the unit interval through the valuation mapping v. Gaines [34] has shown that according to the axiomatic system satisfied by the connectives in $\mathcal{S}$ and the valuation v, one may find a particular representation of $\mathcal{S}$ as 2^U and $v(\alpha) = P([\alpha])$ where $[\alpha]$ is the subset of worlds in U where α is true, and another representation where $\mathcal{S} = [0,1]^U$ and $v(\alpha) = \mu_{[\alpha]}(u)$ where u is fixed in U, and α is a fuzzy statement. Both fuzzy sets and probability can thus be cast in a more general framework of weighted logics. Only in such a general setting does it make sense to compare, say Bellman and Giertz [3] fuzzy set axioms and Kolmogorov axioms. Beyond its mathematical and philosophical interest, Gaines' setting should not however lead us to confuse degrees of probability and membership grades (e.g. Blockley [6], Kovalerchuk and Shapiro [48]).

1.2. Fuzzy Relative Cardinality and Conditional Probability

It F is a fuzzy set on U, its cardinality is $|F| = \sum_u \mu_F(u)$. A well-known index of inclusion of F in another fuzzy set G is (e.g. Dubois and Prade [17])

$$I(F,G) = \frac{|F \cap G|}{|F|} \tag{4}$$

where intersection is defined as minimum. Kosko [47] has pointed out the analogy between I(F,G) and a conditional probability $P(B \mid A)$ where B and G play the same role. He even noticed that the analogy carries over to Bayesian inference, by noticing that

$$I(F,G) \geq \frac{I(G,F) \cdot I(U,G)}{I(G,F) \cdot I(U,G) + I(\bar{G},F) \cdot I(U,\bar{G})} \tag{5}$$

where I(U,G) is nothing but the relative cardinality $\frac{|G|}{|U|}$. Indeed, (5) reminds us of Bayes theorem

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A \mid B) \cdot P(B) + P(A \mid \bar{B}) \cdot P(\bar{B})}. \tag{6}$$

The inequality in (5) is related to $I(F,G) + I(F,\bar{G}) \geq 1$ due to the lack of excluded middle law and the use of min for $\cap$ in (4). The equality holds in (5) when G is a crisp set. It would hold as well with $\cap$ expressed as a product in (4). Kosko [47] claims that (5) implies Bayes theorem when both G and F are crisp. This claim is debatable because in that case (4) is only a special kind of conditional probability, namely $P(B \mid A) = |A \cap B| \;/\; |A|$ provided that P is uniformly distributed on U.

A natural way to generalize both (5) and (6) is to consider a probability measure P on U and $[0,1]^U$ as a set of fuzzy events [82]. Then if we let

$$P(F) = \sum p(u)\mu_F(u) = \int_0^1 P(F_\lambda)d\lambda,$$

where F_λ is the λ-level cut of F (Höhle [41]), be Zadeh's probability of the fuzzy event F, the natural extension of (4) is

$$P(G \mid F) = \frac{P(F \cap G)}{P(F)} \tag{7}$$

and (5) can be generalized changing I(X,Y) into $P(Y \mid X)$. (7) generalizes both a conditional probability and the fuzzy inclusion index advocated by Kosko [47]. Note that a whole literature on conditional probabilities of fuzzy events exists, starting with Okuda et al. [57]. See Dubois, Prade and Yager [27] for a list of basic references.

Kosko [47] contains a derivation of the inclusion index (4) from first principles. He criticizes Cox [10]'s derivation of the standard conditional probability, as arbitrarily assuming that the function f such that $P(A \cap B) = f(P(A \mid B), P(B))$ is twice differentiable. However the unicity of conditional probability only requires that f be *strictly* increasing in both places and continuous, due to Aczel [1] results on the representation of associative real functions.

Concludingly, the comparison between fuzzy relative cardinality and conditional probability should not lead us to conclude with Kosko [47] that "probability is a special case of fuzziness" but leads us to consider probability of fuzzy events, a situation where, as Pearl [60] himself admits, probability and fuzzy sets are orthogonal.

1.3. Possibility Theory is not Compositional

Related to fuzzy sets is the theory of possibility that has been around for 15 years under this name (Zadeh [83], Dubois and Prade [18]) but whose calculation rules go back to Shackle [65] at least. Possibility theory appears as a more direct challenger of probability theory than fuzzy sets because it also proposes a set-function that quantifies the uncertainty of events. A possibility measure on a finite set U is a mapping from 2^U to [0,1] such that

$$\Pi(\emptyset) = 0 \tag{8}$$
$$\Pi(A \cup B) = \max(\Pi(A),\Pi(B)). \tag{9}$$

The condition $\Pi(U) = 1$ is to be added for normal possibility measures. They are completely characterized by a possibility distribution $\pi : U \to [0,1]$ (such that $\pi(u) = 1$ for some $u \in U$, in the normal case) since $\Pi(A) =$

$\max\{\pi(u), u \in A\}$. In the infinite case the equivalence between π and $\prod$ requires that axiom (9) be extended to an infinite family of subsets. Zadeh [83] views the possibility distribution π as determined by the membership function μ_F of a fuzzy set F. However this does not mean that the two concepts of a fuzzy set and of a possibility distribution are the same thing. Zadeh's equation $\pi_x(u) = \mu_F(u)$ is rather similar to equating a likelihood function to a conditional probability (which does not mean that likelihoods are additive measures !). Indeed $\pi_x(u)$ is short for $\pi(x = u \mid F)$ since it estimates the possibility that the variable x is equal to u, knowing the incomplete state of knowledge "x is F", and $\mu_F(u)$ estimates the degree of compatibility of the precise information x = u with the statement to evaluate "x is F".

The controversy between possibility and probability measures is related to the one about fuzzy sets. Cheeseman [7] criticizes possibility theory because he thinks $\prod(A \cap B) = \min(\prod(A),\prod(B))$ holds. But in fact only the inequality $\prod(A \cap B) \le \min(\prod(A),\prod(B))$ universally holds, and part of Cheeseman's objections thus make no sense. But his mistake may be due to the dangerous temptation to consider that possibility measures satisfy a companion axiom to (9) for conjunction as well (since possibility theory is based on fuzzy sets). This is based on a misunderstanding of the kind pointed out in Section 1.1, i.e. confusing membership functions and set functions.

In fact possibility theory is not compositional (contrary to what Pearl [61] p. 679 says). Possibility measures are decomposable with respect to union only, and $N(A) = 1 - \prod(\bar{A})$, where $\bar{A}$ is the complement of A ,is a degree of certainty or necessity which is compositional with respect to intersection only, i.e. $N(A \cap B) = \min(N(A),N(B))$. For instance when one is totally ignorant about event A we have $\prod(A) = \prod(\bar{A}) = 1$ (and $N(A) = N(\bar{A}) = 0$) while $\prod(A \cap \bar{A}) = 0$ and $N(A \cup \bar{A}) = 1$. This ability to model ignorance in a non-biased way is typical of possibility theory, while probability theory in its Bayesian approach, cannot account for ignorance. This is why possibility theory and probability theory, by modelling different kinds of states of knowledge, are complementary theories of uncertainty.

2. BRIDGES

Instead of pursuing sterile polemics, another attitude is to relate fuzzy sets, possibility distributions, and probability. The history of such relationships is already quite long and we shall point out existing bridges between probability and fuzzy sets viewed as possibility distributions.

2.1. Upper Probability

A first point of view is to acknowledge a possibility measure as an upper probability envelope. Namely if we consider a set of probabilities $\mathcal{P}(\prod) = \{P \mid P(A) \le \prod(A), \forall A \subseteq U\}$, then the induced upper probabilities $P^*(A) = \sup\{P(A), P \in \mathcal{P}(\prod)\}$ coincides with possibility degrees $\prod(A)$. The possibility distribution π is then defined by :

$$\pi(u) = P^*(\{u\}), \forall u \in U. \qquad (10)$$

The set $\mathcal{P}(\prod)$ is never empty due to the normalisation of the possibility distribution. This view is adopted in Dubois and Prade [19] for instance. More recently they give a characteristic condition under which a set of probability measures induced by lower bounds of probabilities of specified events lead to possibility measures. Namely $\mathcal{P} = \{P \mid P(A_i) \ge a_i, i = 1,n\} = \mathcal{P}(\prod)$ for some possibility measure $\prod$ if and only if the A_i's are nested (Dubois and Prade [24]). Upper probabilities are closely related to Choquet [8] capacities, and so are possibility measures (e.g. Puri and Ralescu [62]).

2.2. Random Sets and Belief Functions

This view of possibility measures as upper bounds of a family of probability measures has been adopted in the late seventies in a more restricted framework namely that of random sets and that of belief function theory (Shafer [66]). Namely several scholars such as Fortet and Kambouzia [33], Orlov [58], Kampé de Fériet [44], Sales [64], Goodman [38], Nguyen [56], Wang and Sanchez [78], Dubois and Prade [15] have pointed out that given a random set $\mathcal{R}$ on U, i.e. a set of subsets $A_1, \ldots, A_n \subseteq U$, and a mass assignment m such that $m(A_i) = m_i \in (0,1]$ and $\sum_{i=1,n} m_i = 1$, then a membership function could be defined as

$$\mu_F(u) = \sum_{u \in A_i} m_i. \qquad (11)$$

Viewing $\{(A_i,m_i), i = 1,n\}$ as a random set F with realizations A_i, the membership grade $\mu_F(u)$ is then interpreted as the probability that $u \in F$. This kind of probabilistic view of a membership function seems to go back to E. Borel, as pointed out by Godal and Goodman [37]. Interpreting the A_i's as imprecise observations, the two following set functions are defined in belief function theory (Shafer [66])

$$\mathrm{Bel}(A) = \sum_{A_i \subseteq A} m_i \quad ; \quad \mathrm{Pl}(A) = \sum_{A_i \cap A \ne \emptyset} m_i \qquad (12)$$

respectively called belief and plausibility functions. These set functions coincide with lower and upper probabilities respectively and (10) and (11) coincide in this setting due to $\pi(u) = \mathrm{Pl}(\{u\})$. Moreover Shafer [66] showed in his book that $\mathrm{Pl}(A \cup B) = \max(\mathrm{Pl}(A),\mathrm{Pl}(B))$ if and only if the A_i's are nested. Hence a possibility measure is a special case of a plausibility function, itself a special case of upper probability.

2.3. Likelihood Function

Another point of view is to interpret the membership function of a fuzzy set as a likelihood function. This idea is actually quite old in fuzzy set theory (e.g. a Russian 1966 paper by Loginov [52], whose English translation talks about "hussy" sets). It has been the basis of experimental methods for constructing membership functions. Given a population of individuals and a fuzzy concept F, each individual is asked whether a given element $u \in U$ can be called an F or not. The likelihood function $P('F' \mid u)$ is then obtained and represents the proportion of individuals that answered yes to the question. 'F' is then the corresponding non-fuzzy event. Then it is natural to let

$$\mu_F(u) = P('F' \mid u),\ \forall\ u \in U. \tag{13}$$

This type of yes-no experiments have been used by Hersh and Caramazza [39], and Hisdal [40]. Cheeseman [7] exploits this link and claims that on such grounds fuzzy sets are nothing new. See also Stallings [72] for an older but similar critique. Here we suggest that this link will lead to a cross-fertilisation of fuzzy set and likelihood theories, provided that one does not stick to a dogmatic Bayesian position.

Note also that the fundamental point in (13) is to view the likelihood in terms of a conditional uncertainty measure, here a probability. We may think of using other uncertainty measures as well, for instance a possibility measure, i.e.

$$\mu_F(u) = \Pi('F' \mid u),\ \forall\ u \in U.$$

This expresses the equality of the membership function describing the fuzzy class F viewed as a likelihood function with the possibility that an element u is classified in F. This can be justified starting with a possibilistic counterpart of Bayes theorem, i.e.

$$\min(\pi(u \mid 'F'), \Pi('F')) = \min(\Pi('F' \mid u), \pi(u))$$

assuming that no a priori information is available, i.e. $\forall u$, $\pi(u) = 1$, which leads to $\pi(u \mid 'F') = \Pi('F' \mid u)$ (Dubois and Prade [22]).

2.4. Fuzzy Sets in Statistical Inference

Direct relationships between possibility distributions and likelihood functions have been pointed out by Smets [69] and Thomas [74], [75]. Indeed the likelihood function is treated as a possibility distribution in classical statistics for so-called likelihood ratio tests. Namely if some hypothesis of the form $u \in A$ is to be tested against the opposite hypothesis $u \notin A$ on the basis of observation O alone, and the knowledge of elementary likelihood functions $P(O \mid u)$, $u \in U$, then the likelihood ratio test methodology suggests the comparison between $\max_{u \in A} P(O \mid u)$ and $\max_{u \notin A} P(O \mid u)$, i.e. $\Pi(A)$ and $\Pi(\bar{A})$ letting $\pi(u) = P(O \mid u)$ (see e.g. Barnett [2], p. 150). Stallings [72] points out the similarity between maximum likelihood pattern classification and fuzzy pattern classification. Natvig [53] interprets likelihood functions as possibility distributions in the law of total probabilities. Thomas [75] also indicates that the Bayesian updating procedure

$$p(u \mid O) = \frac{P(O \mid u) \cdot p(u)}{P(O)} \tag{14}$$

can be reinterpreted in terms of fuzzy observations. For instance it p represents the base rate of the size of some population, and one learns the concerned individual is "tall", where "tall" is defined by membership function μ_{tall}, then the a posteriori probability can be computed in accordance with (7) as follows

$$p(u \mid tall) = \frac{\mu_{tall}(u) \cdot p(u)}{P(tall)} \tag{15}$$

where $\mu_{tall}(u)$ is interpreted as a likelihood function, and $P(tall) = \int \mu_{tall}(u)\ dP(u)$ is Zadeh [82]'s probability of a fuzzy event (see also Dubois and Prade [22]).

The two probability-oriented views (upper probability and likelihood) of fuzzy sets and possibility distributions are not antagonistic and can be reconciled. For instance (11) corresponds to another experiment for constructing membership functions whereby individuals are asked to point out a single crisp subset $A_i \subseteq U$ that represents some fuzzy concept F best. m_i represents the proportion of individuals for which F is best described by A_i. It makes sense to relate this experiment with the likelihood Yes-No experiment provided that, when each individual chooses A_i as representing F, it means that he would answer yes in the other experiment to the question "is u an F ?" if and only if $u \in A_i$. Then as pointed out in Dubois and Prade [21]

$$P('F' \mid u) = \Sigma_{u \in A_i} m_i. \tag{16}$$

Another way of relating the two approaches is to notice that the use of Dempster rule (Shafer [66]) to combine two belief functions, one of which being a probability function, leads to a probability function that is computed via (14) or (15) exactly. Namely let Bel_1 be defined by $\{(A_i, m_i), i = 1,n\}$ and $Bel_2 = P$, then

$$\forall\ u \in U,\ Bel_1 \oplus Bel_2(u) = \frac{\Sigma_{i:u \in A_i} m_i \cdot p(u)}{\Sigma_{u,i:u \in A_i} m_i \cdot p(u)} = \frac{\mu_F(u) \cdot p(u)}{\Sigma_{u \in U} \mu_F(u) \cdot p(u)}$$

where $Bel_1 \oplus Bel_2$ is a probability measure and μ_F is defined by (11). Only the plausibility of singletons (i.e. the possibility distribution) is useful for the updating, and it plays the role of a likelihood function.The above results suggest how to get likelihood functions from belief functions.

As recalled in Dubois, Moral and Prade [14] the converse problem is addressed by Shafer [66]. In this paper it is suggested that an observation O defines a consonant belief function (whose associated plausibility is a possibility measure) with mass assignment given by

$$m_O(A_i) = P(O \mid u_i) - P(O \mid u_{i+1}) \quad A_i = \{u_1, \ldots, u_i\}$$
$$m_O(U) = P(O \mid u_n)$$
$$m_O(A) = 0 \text{ otherwise}$$

where $U = \{u_1, \ldots, u_n\}$ with $P(O \mid u_1) \geq \ldots \geq P(O \mid u_n)$. This is a non-normalized mass assignment. However if $P(O \mid u_1) = 1$, and letting $\pi(u) = P(O \mid u)$ then this procedure is exactly the converse of equation (11).

Another way of putting together the upper probability approach and the likelihood approach to possibility theory is to consider a family of likelihood functions and to define a possibility as the upper envelope of this family (see Dubois, Moral and Prade [14]).

2.5. Probabilistic Interpretations of Fuzzy Set Operations

Viewing the membership function μ_F as a likelihood function leads to interpret fuzzy set-theoretic connectives in the framework of probability theory. For instance the minimum rule of fuzzy sets is then seen as

$$P('F' \cap 'G' \mid u) = \min(P('F' \mid u), P('G' \mid u)) \qquad (17)$$

where 'F' and 'G' are the crisp events consisting of assigning labels F on G to u respectively. People have then claimed the total dependence between the fuzzy sets F and G, thus pointing out a limitation in fuzzy set theory. This claim is debatable for two reasons : first the crisp events 'F' and 'G' are related to the fuzzy sets F and G only through a random experiment that is built as a model for constructing the membership function. So the dependence is between the crisp events, not the fuzzy sets. This remark also points out that we do not succumb here to the fallacy described in Section 1.1. A second reason is that if we remember the role of possibility distributions as upper bounds the equality in (17) should be turned into an inequality, the right-hand side being an upper bound of the left-hand side. In that perspective the minimum rule expresses a lack of knowledge about dependence (this is the meaning of the non-interactivity assumption in possibility theory) rather than an assumption of strong dependence (see e.g. Dubois, Moral and Prade [14]). The likelihood interpretation of membership functions cannot capture all fuzzy set connectives anyway. Only some connectives below minimum and above maximum are amenable to a likelihood interpretation, because due to inclusion monotonicity $P(A \cap B \mid x) \le \min(P(A \mid x), P(B \mid x))$ and $P(A \cup B \mid x) \ge \max(P(A \mid x), P(B \mid x))$. Hence average like fuzzy connectives and symmetric sums (see Dubois and Prade [18] Chapt. 3 for instance) cannot be captured in this way.

The (consonant) random set view of membership functions also provides a semantics to the main possible choices for defining fuzzy set union and intersection Orlov [58], Goodman [38] (see Dubois and Prade [21] for an overview). But this framework seems to be much more flexible than the one of likelihoods for describing fuzzy set connectives, since in that case fuzzy set intersections or unions correspond to forms of random set intersections or unions, but other connectives such as averages continue to make sense with random sets.

Lastly fuzzy set connectives can also be interpreted in the setting of upper and lower probabilities. In that case again the minimum rule should not be understood as induced by a strong dependence assumption but as a minimal commitment rule when nothing is known about dependence (see Dubois and Prade [24]).

2.6. Possibility Degrees as Infinitesimal Probabilities

Recently Spohn [71] has proposed a theory for the representation of epistemic states that bear strong similarities with possibility and necessity measures, and tries to model the idea of ordering between possible worlds. What he calls an ordinal conditional function comes down to a mapping from a finite set of events to the set of positive integers, denoted κ and such that $\kappa(A \cup B) = \min(\kappa(A),\kappa(B))$; moreover there is an elementary event $\{u\}$ such that $\kappa(\{u\}) = 0$. $\kappa(A)$ expresses a degree of disbelief in A, and grows as A becomes less plausible. It is easy to check that for any real $k > 1$, $1 - k^{-\kappa(\bar{A})}$ is a degree of necessity (see Dubois and Prade [23]). A probabilistic interpretation of $\kappa(A)$ has been suggested by Spohn [71]. Namely, $\kappa(A) = n$ is interpreted as a small probability of the form ε^n, that is the probability of a rare event. Indeed, if A has a small probability with order of magnitude ε^n and B has also a small probability of the form ε^m, then $P(A \cup B)$ is of order of magnitude $\varepsilon^{\min(m,n)}$. These remarks may lead to an interpretation of possibility and necessity measures in terms of probabilities of rare events. Clearly, this relationship between possibility and probability theory is completely unrelated to the ones in the preceding paragraphs. However it explains the strong similarities between Pearl [60]'s approach to non-monotonic reasoning based on infinitesimal probabilities and the possibilistic approach to the same problem (Benferhat et al. [4]).

3. GAPS

Instead of interpreting fuzzy sets and possibility theory in a probabilistic framework one might on the contrary look for interpretive settings for fuzzy sets in which the additive notion of probability is no longer needed. Two such settings are reviewed, that respectively relate possibility to preference and to similarity.

3.1. Possibility as Preference

It is not always meaningful to relate uncertainty to frequency. Some types of events can be rare, unrepeatable, or statistical data may simply be unavailable. It does not prevent people from thinking that some events are more possible, probable or certain than others. And indeed the most primitive way of comparing events is to define a complete pre-ordering on 2^U, that expresses comparative possibility, probability or belief. Comparative counterparts of probability have been searched for since the thirties (see Fishburn [32] for an overview). However no fully satisfying justification of additive probability has been proposed so far, in the finite case.

Comparative possibility is more recent. The complete preordering $\ge_\Pi$ such that $A \ge_\Pi B$ means A is at least as possible as B should satisfy the basic axiom :

$$\forall C,\ A \ge_\Pi B \Rightarrow A \cup C \ge_\Pi B \cup C. \qquad (\Pi)$$

Such comparative possibility relations can be found in Lewis [51]' conditional logics and have been rediscovered by Dubois [13] by suitably modifying the axioms of comparative probability. In that paper it is proved that the only numerical counterparts of comparative possibility are possibility measures. The dual comparative certainty relations ($A \geq_C B \Leftrightarrow \bar{B} \geq_\Pi \bar{A}$) play a crucial role in problems of revision of propositional theories under the name "epistemic entrenchment" (Gärdenfors [35] ; see Dubois and Prade [23]). A comparative possibility relation on 2^U describing the location of an unknown variable x induces a complete preordering on U that can be viewed as a preference relation on the possible values of x. These qualitative possibility distributions can be analyzed from the point of view of their informational content (specificity), as already done for numerical possibility distributions (see, e.g. Benferhat et al. [4], Yager [80]).

3.2. Possibility as Similarity

There is a whole trend in fuzzy set theory according to which the degree of membership $\mu_F(u)$ reflects the similarity between u and an ideal prototype u_F of F (for which $\mu_F(u_F) = 1$). This interpretation of partial membership is clearly related to the relation of distance, and not to probability. Then if a variable x is attached a possibility distribution $\pi = \mu_F$, x = u is all the more possible as u looks like u_F, is close to u_F. This trend is well represented by the so-called "type B membership function" described by Zimmermann [85], and for which direct scaling methods can be applied.

Ruspini [63] has pushed this view further by exploring how the similarity notion R that exists on the referential U affects the inference process between propositions $\alpha, \beta, \ldots$ whose truth or falsity refers to U as a set of possible worlds. For Ruspini, given two classical propositions α and β, the extent to which α implies β is the minimal amount of "stretching" that is required to include the set of worlds in which α is true in the neighborhood of the worlds where β is true. More formally if U is the set of possible worlds, if $u \vDash \alpha$ means α is true in $u \in U$, then the following metric counterpart of Carnap's confirmation degree is obtained

$$I(\beta \mid \alpha) = \inf_{u \vDash \alpha} \sup_{u' \vDash \beta} \mu_R(u,u') \tag{18}$$

where R is a fuzzy proximity relation, that expresses indiscernibility on U. In order to clarify the meaning of $I(\beta \mid \alpha)$, let us consider the case when R is a standard equivalence relation. Let β^* be the upper approximation of β in the sense of rough set theory (Pawlak [59]), i.e. $u \vDash \beta^* \Leftrightarrow \exists u'$, u R u' and $u' \vDash \beta$; then $I(\beta \mid \alpha) = 1$ corresponds to the classical entailment $\alpha \vDash \beta^*$. We get very close to the modal logic of rough sets developed by Fariñas del Cerro and Orlowska [30]. Note that possibility as preference leads to another non-probabilistic degree of entailment namely (e.g. Benferhat et al. [4]) :

$$\begin{aligned} N(\beta \mid \alpha) &= \inf_{u \vDash \bar{\beta} \wedge \alpha} 1 - \pi(u) \text{ if } \Pi(\beta \wedge \alpha) > \Pi(\bar{\beta} \wedge \alpha) \\ &= 0 \text{ otherwise.} \end{aligned} \tag{19}$$

It is positive only if β is true in all the preferred worlds where α is true, i.e. this entailment is preferential in the sense of Shoham [68]. Similarity-based and preference-based entailment degrees are orthogonal with respect to each other, and both contrast with a conditional probability, used by Carnap as a degree of confirmation.

4. POSSIBILITY-PROBABILITY TRANSFORMATIONS

The problem of transforming a possibility distribution into a probability distribution and conversely has received more and more attention. This question is meaningful in the scope of uncertainty combination with heterogeneous sources (some supplying statistical data, other linguistic data, for instance).

However rising the issue means that some consistency exists between possibilistic and probabilistic representations of uncertainty. The basic question is whether it is a mere matter of translation between languages "neither of which is weaker or stronger than the other" (quoting Klir and Parviz [46]). Adopting this assumption leads to transformations that respect a principle of uncertainty and information invariance, as done by Klir [45], on the basis that the entropy H(p) of a probability distribution p and the non-specificity $NS(\pi)$ (or alternative indices of the same kind) of a possibility distribution play the same role in each theory. An uncertainty-invariant transformation between π and p is one such that $H(p) = NS(\pi)$. Klir [45] also requires that they be log-interval scale transformations and satisfy the consistency condition $\pi(u) \geq p(u)$, $\forall u$, stating that what is probable must be possible (Zadeh [83]).

The problem with this approach lies in the assumed equality $H(p) = NS(\pi)$ between quantities whose commensurability sounds debatable. Changing it into proportionality would claim that the entropy ordering of probability exactly reflects the specificity ordering of possibility.

Another point of view on transformations is that possibility and probability theories have distinct roles in describing uncertainty but do not have the same descriptive power. Although it is clear that there are some states of information that probability can describe while possibility cannot (e.g. total randomness) and conversely (a probability distribution cannot express ignorance), it can be advocated that the possibilistic representation of ignorance is weaker than the probabilistic representation, in the sense that the first is additive and the other relies on an ordinal structure induced by the consonance assumption. It does not mean that strict probabilistic representations subsume possibilistic representations, since possibilistic representations can capture weaker states of information that probability distributions cannot. This view is also supported by the mathematical fact that possibility measures coincide with a special class of upper bounds of probabilities. Hence a possibility distribution represents a set of probability measures, i.e. a weaker knowledge than the one of a single probability measure. Klir and Parviz [46] suggest that a probability distribution can be viewed as the lower bound of a family of possibility distributions as well; but it does not prove that a probability *measure* can always be interpreted as a lower possibility *measure*, i.e. that $\forall P$, $\forall A$, we have $P(A) = \inf\{\Pi(A), \Pi \in \mathcal{F}(P)\}$ where $\mathcal{F}(P) =$

$\{\Pi \mid \Pi(A) \geq P(A), \forall A\}$ and indeed $\mathcal{F}$ often possesses a least element, that is a possibility measure itself.

If we accept that possibility distributions are weaker representations of uncertainty than probability distributions, the transformation problem must be stated otherwise. Namely going from possibility to probability leads to increase the informational content of the considered representation, while going the other way around means a loss of information. Hence the principles behind the two transformations are different and we obtained asymmetric transformations (Dubois et al. [26]) :

- From possibility to probability : a generalized Laplacean indifference principle is adopted : the weights m_i bearing on the nested family of levels cuts of π are uniformly distributed on the elements of these level-cuts. This transformation, already proposed by Dubois and Prade [15] consists in selecting the gravity center of the set $\mathcal{P} = \{P \mid \forall A,\ P(A) \leq \Pi(A)\}$ of probability distributions dominated by Π. This transformation also coincides with the so-called pignistic transformation of belief functions (Smets [70]). The rationale behind this transformation is to minimize arbitrariness by preserving the symmetry properties of the representation.
- From probability to possibility : in this case the rationale is to preserve as much information as possible ; hence we select as the result of the transformation of P, the most specific element of the set $\mathcal{F}(P)$ of possibility measures dominating P. See Dubois and Prade [15], and Delgado and Moral [12] for two different characterizations of this transformation.

If we let $\pi(u_i) = \pi_i$ and $p(u_i) = p_i$ and assume $\pi_1 = 1 \geq \ldots \geq \pi_n \geq \pi_{n+1} = 0$ and $p_1 \geq \ldots \geq p_n \geq p_{n+1} = 0$ the two transformations take the following form

$$p_i = \Sigma_{j=i,n} \frac{\pi_j - \pi_{j+1}}{j}$$
$$\pi_i = \Sigma_{j=i,n}\, p_j .$$

As expected they are not the converse of each other.

An interesting question pertaining to possibility-probability transformations is their consistency with respect to second order properties such as marginalization, conditionalisation, independence and the like. Sudkamp [73] recently proved general negative results along this line regarding exact coherence. For instance the projection of a transform is generally different from the transform of the projection. Also, the transform of a product measure is generally not the joint possibility of the transforms of the marginal probabilities. These results should not be surprizing because the probability-possibility transformations cannot be structural homomorphisms due to the strikingly different properties of the basic operations in each calculus. It some consistency is to be looked for it must be a weak form of consistency like the following :

Proposition : If $P \in \mathcal{P}(\Pi)$ then $\mathrm{Proj}_1(P) \in \mathcal{P}(\mathrm{Proj}_1(\Pi))$ where P and Π are defined on $U_1 \times U_2$, and Proj_1 is the marginalization operator on U_1.

The reason why this property holds is that the marginalization process in belief function theory subsumes those of possibility and probability theory (Shafer [67]). In the case of conditionalization and independence the problem is more tricky. First, probabilistic independence is a stronger property than possibilistic non-interaction ; moreover the probabilistic approach to combining possibility distributions leads to the loss of consonance. Hence only consonant approximations of the resulting combination make sense. In the case of conditionalization, several kinds of conditional possibilities exists : one is based on purely ordinal arguments as in equation (19). Hence there is no way of finding consistency with probabilistic arguments. Another conditioning is based on Dempster's rule of conditioning. However the implication $P \in \mathcal{P}(\Pi) \Rightarrow \forall B,\ P(\cdot \mid B) \in \mathcal{P}(\Pi(\cdot \mid B))$ will not hold because $\mathcal{P}(\Pi(\cdot \mid B))$ is generally strictly included in $\{P(\cdot \mid B) \mid P \in \mathcal{P}(\Pi)\}$. Dempster's rule of conditioning does not agree with the upper and lower probability view of belief functions, e.g. Kyburg [50]. The above consistency result with respect to conditionalization will hold if we use the focusing conditional (Dubois and Prade [25]) :

$$\forall B, \Pi(A|B) = \frac{\Pi(A \cap B)}{\Pi(A \cap B) + N(\bar{A} \cap B)} = \sup\{P(\cdot|B) \mid P \in \mathcal{P}(\Pi)\}$$

which is a possibility measure with possibility distribution

$$\pi(u \mid B) = \max\left(\pi(u), \frac{\pi(u)}{\pi(u) + N(B)}\right).$$

This view of conditioning is in accordance with De Campos et al. [11] and Fagin and Halpern [29] for fuzzy measures and belief functions respectively. Clearly much work remains to be done in order to get sufficient insight on the robustness and consistency of possibility-probability transformations with respect to second order properties.

CONCLUSION

This survey was meant to point out that a fuzzy set is not a one-sided object, and that investigations of the relationships between fuzzy set, possibility and probability may be fruitful in the sense that they shed new light on old practice. For instance fuzzy set-theoretic operations can be justified from several probabilistic points of view according to whether the membership function is viewed as an upper probability, the one-point coverage of a random set, or a likelihood function. Conversely the possibilistic nature of likelihood seems to be in accordance with the way statisticians have used them, it leads to use fuzzy set theory as a basis for deriving likelihood functions, and fuzzy events can be considered as relevant evidence in statistics. There are already several domains of application where fuzzy sets and probability are conjointly used. A typical example is fuzzy random variables (e.g. Kruse and Meyer [49], Negoita and Ralescu [55]) that can handle opinion polls where people can give fuzzy (linguistic) answers. Statistics with fuzzy events has received extensive treatment in the past (from Okuda et al. [57]) to Gil [36], and Viertl and Hule [76]. Lastly fuzzy sets can model linguistic probabilities as verified experimentally (Wallsten et al. [77]). Linguistic probabilities can be handled in inference processes, as done by Zadeh [84], Dubois and Prade 20[], Jain and Agogino [43], or in decision analysis as done by Watson et al. [79], and Nau [54] for instance. These examples should convince us that instead of considering probability and fuzzy sets as conflicting rivals, it sounds more promising to build bridges and take advantage of the enlarged framework for modeling uncertainty and vagueness they conjointly bring us to.

REFERENCES

[1] J. Aczél, *Lectures on Functional Equations and Their Applications*. Academic Press, New York, 1966.

[2] V. Barnett, *Comparative Statistical Inference*. Wiley, New York, 1973.

[3] R.E. Bellman, and M. Giertz, "On the analytic formalism of the theory of fuzzy sets," *Information Sciences*, Vol. 5, pp. 149-157, 1973.

[4] S. Benferhat, D. Dubois, and H. Prade, "Representing default rules in possibilistic logic," *Proc. of the 3rd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, Cambridge, Mass., Oct. 26-29, 1992, pp. 673-684

[5] B.M. Bennett, D.D. Hoffman, and P. Murthy, "Lebesgue order on probabilities and some applications to perception," *J. of Mathematical Psychology*, 1992, to appear.

[6] D.I. Blockley, "Fuzziness and probability : a discussion of Gaines' axioms," *Civ. Engng Syst.*, Vol. 2, pp. 195-200, 1985.

[7] P. Cheeseman, "Probabilistic versus fuzzy reasoning," in *Uncertainty in Artificial Intelligence 1*, L. Kanal, and J. Lemmer, Eds., North-Holland, Amsterdam, 1988, pp. 85-102.

[8] G. Choquet, "Theory of capacities," *Ann. Inst. Fourier* (Grenoble), Vol. 5, No. 4, pp. 131-295, 1953.

[9] R. Cooke, *Experts in Uncertainty*. Oxford University Press, Oxford, 1990.

[10] R.T. Cox, "Probability, frequency and reasonable expectation," *American J. of Physics*, Vol. 14, pp. 1-13, 1946.

[11] L.M. De Campos, M.T. Lamata, and S. Moral, "The concept of conditional fuzzy measure," *Int. J. of Intelligent Systems*, Vol. 5, pp. 237-246, 1990.

[12] M. Delgado, and S. Moral, "On the concept of possibility-probability consistency," *Fuzzy Sets and Systems*, Vol. 21, No. 3, pp. 311-318, 1987.

[13] D. Dubois, "Belief structures, possibility theory and decomposable confidence measures on finite sets," *Computers and Artificial Intelligence* (Bratislava), Vol. 5, No. 5, pp. 403-416, 1986.

[14] D. Dubois, S. Moral, and H. Prade, "A semantic for possibility theory based on likelihoods," to appear.

[15] D. Dubois, and H. Prade, "On several representations of an uncertain body of evidence," in *Fuzzy Information and Decision Processes*, M.M. Gupta, and E. Sanchez, Eds., North-Holland, Amsterdam, 1982, pp. 167-181.

[16] D. Dubois, and H. Prade, "A class of fuzzy measures based on triangular norms," *Int. J. of General Systems*, Vol. 8, pp. 43-61, 1982.

[17] D. Dubois, and H. Prade, "A unifying view of comparison indices in a fuzzy set-theoretic framework," in *Fuzzy Sets and Possibility Theory — Recent Developments*, R.R. Yager, Ed., Pergamon Press, 1982, pp. 1-13.

[18] D. Dubois, and H. Prade, *Possibility Theory*. Plenum Press, New York, 1988.

[19] D. Dubois, and H. Prade "Modelling uncertainty and inductive inference," *Acta Psychologica*, Vol. 68, pp. 53-78, 1988.

[20] D. Dubois, and H. Prade, "On fuzzy syllogisms," *Computational Intelligence,* Vol. 4, pp. 171-179, 1988.

[21] D. Dubois, and H. Prade "Fuzzy sets, probability and measurement," *Europ. J. of Operations Research*, Vol. 40, pp. 135-154, 1989.

[22] D. Dubois, and H. Prade "Modelling uncertain and vague knowledge in possibility and evidence theories," in *Uncertainty in Artificial Intelligence 4*, R. Shachter, et al., Eds., North-Holland, Amsterdam, 1990, pp. 303-318.

[23] D. Dubois, and H. Prade, "Epistemic entrenchment and possibilistic logic," *Artificial Intelligence*, Vol. 50, pp. 223-239, 1991

[24] D. Dubois, and H. Prade, "When upper probabilities are possibility measures," *Fuzzy Sets and Systems*, Vol. 49, pp. 65-74, 1992.

[25] D. Dubois, and H. Prade, "Focusing and updating : two concepts of conditioning in the setting of non-additive uncertainty models," (Extended Abstracts) *Proc. of the 6th Inter. Conf. on the Foundations and Applications of Utility, Risk and Decision Theory (FUR VI)*, Cachan, France, June 15-18, 1992, 3 pages.

[26] D. Dubois, H. Prade, and S. Sandri, "On possibility/probability transformations," *Proc. of the 4th Inter. Fuzzy Systems Association (IFSA'91) Congress*, Brussels, 1991, Vol. : Mathematics, pp. 50-53. Extended version to appear in : *Fuzzy Logic : State of the Art*, R. Lowen, Ed., Kluwer Academic Publ., 1992, to appear.

[27] D. Dubois, H. Prade, and R.R. Yager (Eds.) *Readings in Fuzzy Sets and Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1993.

[28] R. Duda, J. Gaschig, and P. Hart, "Model design in the prospector consultant system for mineral exploration," pp. 153-167.

[29] R. Fagin, and J.Y. Halpern, "A new approach to updating beliefs," Research Report RJ 7222, IBM, Research Division, San Jose, CA, 1989.

[30] L. Fariñas del Cerro, and E. Orlowska, "DAL — A logic for data analysis," *Theoretical Computer Science*, Vol. 36, pp. 251-264, 1985.

[31] T.L. Fine, *Theories of Probability*. Academic Press, New York, 1973.

[32] P.C. Fishburn, "The axioms of subjective probability," *Statistical Science*, Vol. 1, pp. 335-358, 1986.

[33] R. Fortet, and M. Kambouzia, "Ensembles aléatoires et ensembles flous," *Publications Econométriques*, Vol. IX, No. 1, pp. 1-23, 1976.

[34] B.R. Gaines, "Fuzzy and probability uncertainty logics," *Information and Control*, Vol. 38, pp. 154-169, 1978.

[35] P. Gärdenfors, *Knowledge in Flux — Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, 1988.

[36] M.A. Gil, "On the loss of information due to fuzziness in experimental observations," *Ann. Inst. Statist. Math.*, Vol. 4, No. 4, pp. 627-639, 1988.

[37] R.C. Godal, and T.J. Goodman, "Fuzzy sets and Borel," *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 10, No. 10, p. 637, 1980.

[38] I.R. Goodman, "Fuzzy sets as equivalence classes of random sets," in *Fuzzy Sets and Possibility Theory*, R.

Yager, Ed., Pergamon Press, Oxford, 1982, pp. 327-342.

[39] H.M. Hersh, and A. Caramazza, "A fuzzy set approach to modifiers and vagueness in natural language," *J. of Exp. Psycho., General*, Vol. 105, pp. 254-276, 1976.

[40] E. Hisdal, "Naturalized logic and chain sets," *Information Sciences*, Vol. 57-58, pp. 31-77, 1991.

[41] U. Höhle, "MaBe auf unscharfen Mengen," *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete*, Vol. 36, pp. 179-188, 1976.

[42] E.J. Horvitz, D.E. Heckerman, and C.P. Langlotz, "A framework for comparing alternative formalisms for plausible reasoning," *Proc. of the 5th National Conf. on Artificial Intelligence (AAAI'86)*, Philadelphia, PA, Aug. 11-15, 1986, Vol. 1 : Science, pp. 210-214.

[43] P. Jain, and A.M. Agogino, "Stochastic sensitivity analysis using fuzzy influence diagrams," in *Uncertainty in Artificial Intelligence 4*, R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, Eds., North-Holland, Amsterdam, pp. 79-92, 1990.

[44] J. Kampé de Fériet, "Interpretation of membership functions of fuzzy sets in terms of plausibility and belief," in *Fuzzy Information and Decision Processes*, M.M. Gupta, and E. Sanchez, Eds., North-Holland, Amsterdam, 1982, pp. 93-98.

[45] G.J. Klir, "A principle of uncertainty and information invariance," *Int. J. of General Systems*, Vol. 17, No. 2-3, pp. 249-275, 1990.

[46] G.J. Klir, and B. Parviz, "Probability-possibility transformations : a comparison," *Int. J. of General Systems*, Vol. 21, pp. 291-310, 1992.

[47] B. Kosko, "Fuzziness vs. probability," *Int. J. of General Systems*, Vol. 17, pp. 211-240, 1990.

[48] B.Y. Kovalerchuk, and D.I. Shapiro, "On the relation of the probability theory and the fuzzy setz theory foundations," *Computers and Artificial Intelligence*, Vol. 7, No. 5, pp. 385-396, 1988.

[49] R. Kruse, and K.D. Meyer, *Statistics with Vague Data*. D. Reidel, Dordrecht, 1987.

[50] H.E. Kyburg, Jr., "Bayesian and non-Bayesian evidential updating," *Artificial Intelligence*, Vol. 31, pp. 271-293, 1987.

[51] D. Lewis, *Counterfactuals*. Basil Blackwell, Oxford, 1973. 2nd edition, Billing and Sons Ltd., Worcester, UK, 1986.

[52] V.I. Loginov, "Probability treatment of Zadeh membership functions and their use in pattern recognition," *Eng. Cyber.*, pp. 68-69, 1966.

[53] B. Natvig, "Possibility versus probability," *Fuzzy Sets and Systems*, Vol. 10, pp. 31-36, 1983.

[54] R.F. Nau, "Decision analysis with indeterminate or incoherent probabilities," *Annals of Operations Research — Vol : Choice Under Uncertainty*, P.C. Fishburn, and I.H. LaValle, Eds., 1992, to appear.

[55] C.V. Negoita, and D. Ralescu, *Simulation, Knowledge-Based Computing, and Fuzzy Statistics*. Van Nostrand Reinhold Comp., New York, 1987.

[56] H.T. Nguyen, "On conditional possibility distributions," *Fuzzy Sets and Systems*, Vol. 2, No. 1, pp. 53-66, 1979.

[57] T. Okuda, H. Tanaka, and K. Asai, "A formulation of fuzzy decision problems with fuzzy information, using probability measures of fuzzy events," *Information and Control*, Vol. 38, pp. 135-147, 1978.

[58] A.I. Orlov, "Fuzzy and random sets," in *Prikladnoï Mnogomiernii Statisticheskii Analyz*, Nauka, Moscow, 1978, pp. 262-280 (in Russian).

[59] Z. Pawlak, *Rough sets — Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publ., Dordrecht, 1991.

[60] J. Pearl, *Probabilistic Reasoning Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[61] J. Pearl, "Integrating probability and logic," in *Readings in Uncertain Reasoning*, G. Shafer, and J. Pearl, Eds., Morgan Kaufmann, San Mateo, CA, 1990, pp. 677-679.

[62] M.L. Puri, and D. Ralescu, "A possibility measure is not a fuzzy measure," *Fuzzy Sets and Systems*, Vol. 7, pp. 311-313, 1982.

[63] E.H. Ruspini, "Approximate reasoning : past, present, future," *Information Sciences*, Vol. 57-58, pp. 297-317, 1991.

[64] T. Sales, "Fuzzy sets as set classes," *Stochastica*, Vol. 6, pp. 249-264, 1982.

[65] G.L.S. Shackle, *Decision, Order and Time in Human Affairs*. Cambridge University Press, UK, 1963, (62 pages).

[66] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.

[67] G. Shafer, "Belief functions and possibility measures," in *Analysis of Fuzzy Information — Vol. I: Mathematics and Logic*, J.C. Bezdek, Ed., CRC Press, Boca Raton, Fl., 1987, pp. 51-84.

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

[69] P. Smets, "Possibilistic inference from statistical data," *Proc. of the 2nd World Conf. on Mathematics at the Service of Man*, Las Palmas (Canary Island), Spain, June 28-July 3, 1982, pp. 611-613.

[70] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," in *Uncertainty in Artificial Intelligence 5*, M. Henrion, et al., Eds., North-Holland, Amsterdam, 1990, pp. 29-39.

[71] W. Spohn, "A general non-probabilistic theory of inductive reasoning," in *Uncertainty in Artificial Intelligence 4*, R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, Eds., North-Holland, Amsterdam, 1990, pp. 149-158.

[72] W. Stallings, "Fuzzy set theory versus Bayesian statistics," *IEEE Trans. on Systems, Man and Cybernetics*, pp. 216-219, 1977.

[73] T. Sudkamp, "On probability-possibility transformations," *Fuzzy Sets and Systems*, Vol. 51, pp. 73-81, 1992.

[74] S.F. Thomas, "A theory of semantics and possible inference with application to decision analysis," PhD Thesis, University of Toronto, Canada, 1979.

[75] S.F. Thomas, "Possibilistic uncertainty and statistical inference," *ORSA/TIMS Meeting*, Houston, Texas, 1981.

[76] R. Viertl, and H. Hule, "On Bayes' theorems for fuzzy data," *Statistical Papers*, Vol. 32, pp. 115-122, 1991.

[77] T.S. Wallsten, D.V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth, "Measuring the vague meanings of probability terms," *J. of Experimental Psychology : General,* Vol. 115, No. 4, pp. 348-365, 1986.

[78] P.Z. Wang, and E. Sanchez, "Treating a fuzzy subset as a projectable random set," in *Fuzzy Information and Decision Processes*, M.M. Gupta, and E. Sanchez, Eds., North-Holland, Amsterdam, 1982, pp. 213-220.

[79] S.R. Watson, J.J. Weiss, and M. Donnell, "Fuzzy decision analysis," *IEEE Trans. on Systems, Man and Cybernetics,* Vol. 9, No. 1, pp. 1-9, 1979.

[80] R.R. Yager, "On the completion of qualitative possibility measures," Tech. Report #MII-1233, Machine Intelligence Institute, Iona College, New Rochelle, NY, 1992.

[81] L.A. Zadeh, "Fuzzy sets," *Information and Control*, Vol. 8, pp. 338-353, 1965.

[82] L.A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, Vol. 23, pp. 421-427, 1968.

[83] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems,* Vol. 1, pp. 3-28, 1978. Reprinted in *Fuzzy Sets and Applications : Selected Papers by L.A. Zadeh,* Yager, R.R., S. Ovchinnikov, R.M. Tong, and H.T. Nguyen, Eds, Wiley, New York, 1987, pp. 193-218.

[84] L.A. Zadeh, "Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions," *IEEE Trans. on Systems, Man and Cybernetics,* Vol. 15, pp. 754-763, 1985.

[85] H.J. Zimmermann, *Fuzzy Sets, Decision Making, and Expert Systems*. Kluwer Academic Publ., Boston, 1987.